Appendix

A. Related Literature on Children's Language Development

This section presents a detailed literature review on children's language development. Gough and Turner (1986) proposed an influential theoretical model of how children become proficient readers. According to their model – known as the Simple View of Reading – reading proficiency requires two different skills: decoding skills and language comprehension skills. Decoding, or word recognition, is the capacity to recognize print and read words written on a page. Language comprehension is the ability to make sense of the language one hears or reads. As Scarborough, Neuman, and Dickinson (2009) argued, a child who cannot read proficiently is a child who lacks one or both of these two different skills. However, empirical work suggests that failure to read proficiently is disproportionately due to a lack of language comprehension skills. Foorman et al. (2018) provide compelling evidence about the importance of language proficiency in explaining variability in reading proficiency. The authors quantified language proficiency with listening comprehension[19], vocabulary[20], and syntax measures.[21] The team assessed decoding skills with assessments of phonological awareness[22] and tests of decoding fluency.[23] According to their findings, 80% of reading skills variability at Grade 10 is due to variability in language comprehension skills.[24] The PIAAC study by the OECD (2016) measured literacy proficiency in how well the test takers performed in the test (percent of correct answers) and how fast they answered the questions. Their data show that proficient readers (those at Level 5) had only slightly better performance than below-basic readers (the ones below Level 1) in tests that measured decoding skills. However, below-basic readers had much worse performance and much lower speed in items that measured paragraph interpretation or

[19] The measures of listening comprehension varied across grades. In Grades 1 and 2, the examiner read two narrative passages from the Florida Assessment for Instruction in Reading (FAIR, Florida Department of Education, 2009-2014) and asked the student to retell the passages. In Grade 3, the authors measured listening comprehension with the Clinical Evaluation of Language Fundamentals (CELF-4, Paslawski, 2005).

[20] The team measured expressive vocabulary with FAIR Vocabulary Task (Grades 1 and 2), Peabody Picture Vocabulary Test-4 (PPVT-4, Dunn & Dunn, 2007) and the Study Aid and Reading Assistant (SARA, Sabatini et al., 2013).

[21] The authors measured receptive and expressive syntax with CELF-4.

[22] The study team measured phonological awareness in Grades 1 and 2 with the Comprehensive Test of Phonological Processing-2 (CTOPP-2, Wagner et al., 2012).

[23] The Test of Word Reading Efficiency–2 (TOWRE-2, Torgesen et at., 2011) is the instrument the team used to measure decoding fluency.

[24] In elementary grades, language comprehension unique factors as well as common factors of decoding and language comprehension skills explain all of the variability in reading.

short logical sentence comprehension. In sum, the socioeconomic gaps in language skills substantially predict the socioeconomic gaps in reading proficiency.

Research suggests that early language environment influences language development. Weisleder and Fernald (2013) show that the quality of a child's language environment[25] at baseline (age 19 months) predicts vocabulary and language processing skills[26] five months later, even after controlling for vocabulary and language processing skills at baseline. These findings are consistent with Pan et al. (2005), who showed that the more complex the maternal linguistic input, the larger the child's vocabulary.[27] Gilkerson et al. (2018) report the results of a longitudinal study that first measures the quality of infants' and toddlers' language environment[28] and then, when the same children are nine to fourteen years old, measures their IQ and language development.[29] The authors show that conversational turns between adults and children accounted for 14% to 27% of the later human capital dimensions variance. Romeo et al. (2018) show that children who experienced more conversational turns with adults also exhibited greater activation of an area of the brain linked to language processing[30], as measured by a storytelling task in the fMRI, even after the authors control for the family's socio-economic characteristics and the child's IQ.

Research shows that there is a sizeable socioeconomic gradient in the language environment. The relation between family SES and the child's early language skills is partly due to the quantity and quality of parent speech directed towards the child during day-to-day interactions. For example, Hart and Risley (1995) estimated that high-SES children heard approximately 2,153 words per hour. In contrast, children from low SES families heard only about 616 per hour. Hoff (2003) showed that high-SES mothers use longer utterances and more

[25] The authors measure the child's language environment by the number of words in child-directed speech (that is, not counting speech that is overheard, but not directed to, the child) during a ten-hour period.
[26] The authors use the Looking-While-Listening task (LWL, Fernald et al., 2008) to measure language processing skills. In the LWL procedure, infants look at pairs of pictures while listening to speech naming one of the pictures, and their gaze patterns are video-recorded as the sentence unfolds in time. Reaction time (i.e., the amount of time the infant shifts away from the distracter to the target picture) is the measure of language processing skills. Children with higher processing skills take less time to shift away to the target picture.
[27] See also Hoff (2003).
[28] The quality of the language environment is estimated by the number of conversational turns between adults and children and the number of adult words spoken around the child. These are the same measures we use in our study.
[29] IQ is measured with the Wechsler Intelligence Scale for Children (Wechsler, 2014). Language development is measured with the PPVT and the Expressive Vocabulary Test (Williams, 2007).
[30] Broca's area.

vocabulary words with their children than low-SES mothers.[31] Rowe and Goldin-Meadow (2009) found that high-SES mothers gesture more about objects in the environment when they are close to their infants and toddlers. They also report that lower SES mothers talk less often, use smaller vocabulary, and employ syntactic structures that are less varied or complex. The low-SES mothers also tend to command instead of engage with their children. Higher SES children are more likely to be exposed to rarer vocabulary, more linear narratives, more open-ended questions, and other speech characteristics more closely aligned to the school system's academic language environment (e.g., Fernald et al., 2012; Hart & Risley, 1995; Hoff, 2006).

All in all, it is a desirable goal to have children become proficient readers. Still, many low-income children fail to reach even basic literacy levels because they do not have appropriate language skills. Developing these skills requires greater exposure to language so that children can practice and hone their language processing skills and simultaneously increase their vocabulary.

B. The Context and The Philadelphia Human Development Study

The Philadelphia Human Development (PHD) Study is a longitudinal study that measures parental beliefs about the impact of early investments on early human capital formation. The study recruited 822 English-speaking women in the second trimester of their first pregnancy. Approximately 80% of the participants received prenatal care from inner-city clinics that predominantly served Medicaid-insured patients. The remaining 20% of the study participants received prenatal care from suburban clinics that primarily served privately-insured patients.

The original study design involved three visits. At the time of enrollment, the first visit was at the prenatal-care clinic. We measured parental beliefs and collected data about the study participants' demographic characteristics on the first visit. Table B.1 presents the summary statistics of the two groups of PHD Study Participants (inner-city and suburban samples). The pregnant women in the Medicaid clinics are more likely to be poor, single, and have lower educational attainment. Furthermore, they are less likely to be non-Hispanic white.

The second visit occurred when the baby was between 8-12 months old at the child's home. For the second visit, we encountered 687 of the 822 participants. We used two instruments to measure investments in the early human capital of children. First, in all

[31] See also Huttenlocher et al. (2007). All of these studies count words directed to the child and do not count words the child overhears in adult speech.

households, we assessed the quality of parent-child interactions by the Home Observation for the Measurement of the Environment or HOME (Bradley & Caldwell, 1980). In Table B1, we show that the gap in the HOME score between the inner-city and suburban samples is about 83% of a standard deviation. Cunha et al. (2022) combined data from the first and second rounds of the PHD study and found that parental beliefs elicited in the first round predicted heterogeneity in early investments in children as measured by the HOME inventory in the second visit (about one year later).

Second, we selected 272 families to study the language environment's quality using the LENA Pro System (Gilkerson & Richards, 2008). We obtained valid data for 239 families. Therefore, we recorded the quality of the early language environment that low-income children experience on a typical day when they were about one year old. Table B1 shows that the children in the inner-city family had approximately 306 conversational turns with an adult during a 12-hour recording period, and children in the suburban sample had around 344 conversational turns during the same period. The difference, however, was only marginally statistically significant at the 10% level. It was surprising to see differences in conversational turns at such a young age because children do not yet "talk" (even though they vocalize). Conversational turns matter because show that the differences in the number of conversations initiated by children in high and low socioeconomic groups mainly determine the deficits in the parental linguistic input (words addressed to the child) between groups. We did not find differences in adult word counts in the PHD study. However, as we explain below, the automatic counts include both words used in child-directed speech and speech not directed to the child but captured by the device because the adult is close to the child. Finally, the Inner-City sample children watched one extra hour of TV daily.

The third visit happened at the Children's Hospital of Philadelphia research building when the child was between twenty-two and twenty-six months old. We assessed the development of 674 (out of the 822) children using the Bayley Scale of Infant Development (BSID III). The BSID III contains two language development dimensions: receptive language (ability to understand language input) and expressive language (ability to produce language output). First, we average these two scores to estimate a composite language score. Then, we use the BSID III normalizing statistics to calculate the age-normed score. Table B1 shows that the children in the inner-city sample are 56% of a standard deviation below the mean in language development. In comparison, the children in the suburban sample are 21% of a

standard deviation above the mean. Therefore, we estimate a difference of nearly 77% of a standard deviation in language development around age two years.

A diverse body of literature in the social sciences has extensively documented the differences between socioeconomic groups in investment and human capital (e.g., Duncan and Murnane, 2011). The data from the PHD Study showed that the within-group differences are even more considerable. Table B2 reports our estimates of between and within variation in investment and child development measures. The within-group variation is 89% to 86% of the total variance for the HOME and the Bayley Language Composite Score, respectively. For conversational turns, the variation was almost entirely within groups. A possible result of the difference between the HOME scale and the LENA system is that the former aims to capture a child's environment's permanent aspects. In contrast, the LENA system may record a mixture of permanent and transitory elements of the child's language environment, and transitory shocks vary greatly within groups but little across groups. [32]

Another possibility is that the HOME scale captures the child environment's elements more strongly impacted by household income. At the same time, the LENA system is sensitive to parental behavior that is not influenced by family income. To investigate this issue, we regress standardized HOME and standardized total conversational turn counts against family income quartiles.[33] Indeed, as we present in Table B3, the correlation between family income is stronger with HOME scores and much weaker with LENA scores.

Next, we correlated early investment measures, which we collected on the second visit, with the language development measures we obtained on the third visit (about one year later). Our goal was to compare the predictive performance of the HOME with that of the LENA measures. Therefore, we constrain our analysis with the smaller sample (N = 223) for which we have both types of measures of language environment. We display the results in Table B4. First, we estimate the correlation between our language development assessments with each language environment measure separately. Then, for each of these one-to-one relationships, we estimate four models. The first model does not control for any observable characteristics of

[32] Sampling variability is a challenge of using audio data obtained through one recording day. A child's language environment can fluctuate daily because of variations in adults' or child's moods, the number and personalities of adults in the household, and the child's health (an ill child may sleep more hours, for example). One could reduce sampling variability by recording the audio environment for three or four days and then averaging across days. However, this procedure burdens participant families and requires appropriate funding because of the logistical costs of such an operation.
[33] We adjust for the recording duration when we include the total counts of conversational turns in the model.

the family (but controls for recording duration for the LENA measures). The second model adds demographic covariates (race, ethnicity, dummy for maternal year of birth between 1978-1987, dummy for college education). The third model adds a dummy for the inner-city sample. The fourth model has dummies for quartiles of family income. Panel A shows the results for the HOME, and we can see that the HOME at nine months predicts language development at age 24 months, but the strength of the prediction decreases by almost 73% from Model 1 to Model 4. Panel B displays the results for conversational turn counts. The relationship with language development at age 24 months is stable as we move from Model 1 to Model 4. Panels C and D present the same analysis for adult word counts and exposure to TV, respectively. Once we control for observed characteristics, the one-to-one relationships with language development are weaker and not statistically significant.

Panel E combines the HOME with at least one LENA measure. All models in Panel E have controls for the families' observed characteristics. When we combine the HOME with any measure of LENA, the HOME's coefficient is small, and it is only statistically significant (at the 10% level) when combined with exposure to TV. In contrast, when we coalesce conversational turn counts with the HOME, the latter is not statistically significant, but the former is. Even when considering all four measures together, conversational turns had the largest point estimate, and it was the only one with statistical significance (at 10%).

The language development gap between the inner-city and suburban samples was around 78% of a standard deviation. If we use the smaller LENA System sample in our analysis, the gap is more extensive (approximately 84% of a standard deviation). A one-standard-deviation shift in HOME scores predicts a change of around 13% of a standard deviation in language development (Columns 3 and 4).

Therefore, conversational turns' prediction power was stable and more or less orthogonal to including family characteristics or other measures of parent-child interactions. On the other hand, suppose the automatic counts of conversational turns produced by the LENA System data contain sampling error due to temporary variation and that the sampling error is classical. Suppose, in addition, that the temporary variation is less critical for language development. In that case, our estimates in Table B4 are a lower bound for the language environment's contribution to predicting future language development.

In summary, many (but not all) children growing up in low-income households have deficient language development, and the deficits in language development correlate with gaps in the language environment measured one year earlier. Furthermore, in the PHD Study data,

the language environment's heterogeneity – measured by conversational turn counts – has a fragile association with family income. For these reasons, our study hypothesizes that parental beliefs about the importance of the early language environment for a child's language development drive parental linguistic input. Through the LENA Program, we investigate if it is possible to change parental linguistic input and, if so, whether parental beliefs are one of the mechanisms of this change.

C. Results of the Recruitment for the LENA Start Program

This section describes more results of our randomization procedure and recruitment efforts. Table C1 presents the results of the analysis of our recruitment efforts, consent to participate in the study, and attendance (conditional on the invitation to the LENA Start Program). Let $d_i^F = 1$ if our team found and contacted an eligible study participant and $d_i^F = 0$, otherwise. We estimate a linear probability model in which d_i^F is the dependent variable, and the random assignment to the invitation or control arm is the only independent variable. Column 1 in Table C1 shows that the assignment to control or invitation arms does not predict our team's success in locating a study participant. However, it also shows that our team was more likely to locate older, white, and lower-income mothers within the group of eligible study participants.

Next, we describe the results of the consenting procedures. Let $d_i^C = 1$ if our team successfully consented an eligible study participant, conditional on being located, and $d_i^C = 0$, otherwise. Again, we estimate a linear probability model in which d_i^C is the dependent variable and the random assignment is the independent variable. Table C1 shows that the families randomly assigned to the LENA Start Program invitation arm were less likely to consent to participate in the study. This finding is not unusual in the literature (e.g., see Kalil, 2014) and indicates that such programs have difficulty attracting parents. However, when we do not condition on being located, the random assignment variable does not predict consent.

Third, we discuss the attendance results at the LENA Start Program. Similar to previous exercises, let $d_i^A = 1$ if our team successfully made an eligible study participant to attend at least one session, conditional on consenting to participate in the study, and $d_i^A = 0$, otherwise. Again, we estimate a linear probability model in which d_i^A is the dependent variable, and the random assignment is the only independent variable. Table C1 shows that the random assignment to control or invitation arms was the most crucial variable in predicting whether a family attended at least one of the LENA Start Program sessions. Families randomly assigned to the LENA Start Program invitation arm were likelier to attend the sessions. This result also

holds when we do not condition on having consented. It indicates that the random assignment is not a weak Instrumental Variable.

Finally, we also look at the results for consent and attendance when we use the whole sample (i.e., the 289 eligible observations) in Columns (4) and (5). Column (4) shows that conditional on eligibility, consenting to participate in the program does not show significant differences across treatment and control groups. The same result is detected in Column (5) when we analyze the data on attendance conditional on eligibility.

We finish this section by showing that the randomization to control or invitation group is a strong instrumental variable for attendance to the LENA Start Program (see Table C2).

D. Results on Construction of the Recording Data

Unfortunately, not all parents follow the recording protocol as instructed. We adopted the following criteria to determine whether the data we received from the parents was valid or not valid. We divided each recording session data into five-minute segments. In our dataset, there are 68,407 such segments.[34] A segment is defined to be valid if it satisfies four conditions. First, the segment was complete, meaning the recording lasted precisely 300 seconds. About 1.6% of our recording segments were incomplete, and we dropped them from our final recording dataset.

Second, the segment does not have either of two recording errors. The first recording error arises when the audio file does not have enough child speech. The second type of error occurs when the audio file does not have enough overall speech. Approximately 11% of the recording segments had at least one of the two recording errors, so we dropped them from our analysis.

Third, the recording segment took place between 8:00 AM and 8:00 PM. The objective of imposing this restriction is to improve comparability across families as children differ when they go to bed and wake up. We instructed families not to start the recording session until the child was awake and removed and turned off the device when they went to bed.

Fourth, we required that the recording session last at least two hours. Four families did not provide a valid file with at least two hours of recording. The average recording duration was over fourteen hours, with a standard deviation of six hours.

[34] Included recording segments are not necessarily contiguous. For example, a parent may start the recording session on Saturday at 1 PM, turn off the device when the child goes to bed, and resume the recording on the next day (or even later in the week).

Our team obtained valid recording data from 114 parents for the baseline and 104 for the endline. A total of 90 parents provided valid recording data for both baseline and endline recording sessions. Table D1 presents our analysis of the relationship between adherence to research protocol and assignment to the control or invitation group. We analyze the relationship between the recording data's validity and random assignment for each round. First, we consider the dummy variable equal to one if the family submits a valid recording and zero otherwise and run a probit model. Columns (1) and (4) in Table D1 present the baseline and endline data results. We find no relationship between random assignment to control and invitation arm and submission of valid recording data for baseline or endline. While we expected this finding for the baseline, we could not guarantee this result for the endline because parents who participated in the LENA Start Program became more used to providing valid recording data as they had to record their children's language environment for thirteen weeks to get feedback.[35]

Second, we run an OLS regression on the standardized recording duration. As shown in Columns (2) and (5), the random assignment to control or invitation arms does not predict the recording session's length at baseline, but we find evidence that it does so for the endline session. Families randomly assigned to the invitation arm tend to provide recording files whose duration is 40% of a standard deviation longer. Once we account for selection in Columns (3) and (6), the coefficient for the endline results gives the exact point estimate virtually. Given the large discrepancy in recording length between the study's control and invitation arms, we control for recording length in our analyses. In Table F3, we also consider models with polynomials in recording length as control variables. We show that the differences in recording length between control and invitation arms do not drive our results.

Table D2 compares differences in baseline conversational turn counts between the two groups. Panel A reports the OLS estimators (and respective standard errors) of coefficients β_1 and β_3 for variations of the following regression model:

$$Y_{i,0} = \beta_0 + \beta_1 Z_i + X_i \beta_2 + \beta_3 R_{i,0} + \sum_{j=1}^{J} \gamma_j B_{i,j} + \epsilon_{i,0}. \quad (6)$$

The dependent variable $Y_{i,0}$ represents conversational turn counts measured at baseline, so that $t = 0$, by family i. The binary variable denotes the random assignment to control $(Z_i = 0)$ or invitation arm of the study $(Z_i = 1)$. The vector X_i captures demographic

[35] We do not find evidence that participation in the measurement study increases the chance of submitting valid recording data.

characteristics of family i. The variable $R_{i,0}$ denotes the vector that contains information about the recording at baseline, which includes recording duration and dummy variables for recordings on Saturday or Sunday, respectively. The variable $B_{i,j}$ takes the value one if family i was a member of the randomization group j, and zero otherwise.

The findings are the same across the four models. Minor differences exist in the number of conversational turns between the study's control and invitation arms at baseline. Although the point estimates are not statistically significant, the children's language environment in the study's invitation arm has slightly lower quality. The results also show that the longer the recording duration, the higher the dependent variables' levels.

We also investigate whether there are pre-existing differences between parents who accept the invitation and participate in the LENA Start Program. The set of parents who attend at least one of the LENA Start Program sessions is a potentially selected set of invited parents. For this reason, we instrument attendance to the LENA Start Program with random assignment to control or invitation arms. Let D_i take the value one if the parent attends at least one of the LENA Start Program sessions and zero otherwise. We estimate the following model via Two-Stage Least Squares:

$$D_i = \alpha_0 + \alpha_1 Z_i + X_i \alpha_2 + \alpha_3 R_{i,0} + \sum_{j=1}^{J} \pi_j B_{i,j} + \omega_{i,0} \quad (7)$$

$$Y_{i,0} = \beta_0 + \beta_1 D_i + X_i \beta_5 + \beta_3 R_{i,0} + \sum_{j=1}^{J} \gamma_j G_{i,j} + \epsilon_{i,0} \quad (8)$$

We argue that the random assignment is valid because it satisfies both the exclusion restriction and a solid first stage (see Table C1). We find that the 2SLS estimator of β_1 is not statistically significant for all models.[36] However, our small sample size may drive this statistical insignificance because the point estimates are not close to zero. The statistical insignificance arises because the standard errors are relatively large. Because of the pre-existing differences in the outcome variables, our models to estimate the program's impact consider these pre-existing differences.

E. Additional Details on Survey Measures

[36] Note that when using the clustered standard errors, the 2SLS estimator of β_1 becomes statistically significant (at the ten percent level) for the most comprehensive specification - Model 3

In this appendix section, we describe the operationalization of the data to measure mechanisms.

Self-Efficacy and Sense of Social Support Scales: We used the items in the LENA Start enrollment form (which all LENA Start sites use) to measure parental self-efficacy and parental sense of social support. Each one of these scales has four items. The four items in the self-efficacy scale are:

1. *I have the skills to be the best parent I can be*.
2. *My child will do very well in school*.
3. *I know what my child should be able to do at each age as they grow*.
4. *When my child is upset, I can easily calm him/her down*.

For each one of these items, parents choose one alternative among five. The alternatives range from “least sure” (1) to “very sure” (5).

The four items in the sense of social support scale are:

1. *I am relaxed most of the time when I’m with my baby*.
2. *My family spends a lot of time together*.
3. *It’s easy for me to talk with other parents about being a parent*.
4. *It’s easy for me to ask other parents for help or advice if I need to*.

For each one of these items, parents choose one alternative among five. The alternatives range from “strongly disagree” (1) to “strongly agree” (5).

We show the results of our estimates of the PCM in Table E2. Table E3 examines the balance between control and treatment arms at baseline.

Parental Knowledge Scale: We reproduce the procedures in Suskind et al. (2016). We score each response with a binary variable (0/1), in which “0” and “1” represent, respectively, a wrong and a correct answer. Then, we estimate the average for each topic and the overall average. The results in Table 4 use the overall average, and the results in Table E4 decompose the analysis by topic (so we use the topic-specific average).

Table E3 examines the balance between control and treatment arms at baseline. Table E4 shows that the impact in the aggregate scale masks heterogeneity in the subscales. When we disaggregate the scales, the point estimate of the ITT suggests relatively larger impacts on “Reading Books,” “Learning Math,” and the “School Readiness” subscales and more negligible impacts on “Learning to Talk” or “Language and TV” subscales. The point estimate of the LATE

parameter suggests impacts on “Learning Math” and the “School Readiness” subscales but no impacts on the other three scales.

Parental Beliefs: We ask parents the following question:

Imagine a two-year-old child who is average in terms of language development. Consider the following two scenarios. Scenario 1: the adults in the home talk a lot to the child and often read books to the child, but the child does not watch a lot of shows for kids (for example, Sesame Street) on TV. Scenario 2: the adults in the home do not talk a lot to the child and rarely read books to the child, but the child watches a lot of shows for kids (for example, Sesame Street) on TV. What do you think will the child's language development be when the child is three years old?

We give parents five alternatives: Low, low-average, average, high-average, and high. To produce our estimates, we make two assumptions.

First, we map these alternatives to percentiles in the distribution of language development: 5th, 25th, 50th, 75th, and 95th percentiles, respectively.

Second, we assume that the distribution of language development at age three years is normal with mean zero and variance one. We then use the Z scores associated with the percentiles. We take parental beliefs as the difference between the Z scores for “Scenario 1” and the Z scores for “Scenario 2.” To ensure that the extreme percentiles do not drive our results, we replace the 5th and 95th percentiles with the 10th and 90th percentiles, respectively. We show the results in Table E1 and note that our findings are robust to these choices.

F. Robustness Analysis

This section compares our estimated treatment effects with the existing literature. We also show results using alternative methods, including the fixed-effect (FE) model and an instrumental variable fixed-effect (IV-FE) model. Finally, we also present the findings of other sensitivity checks.

First, to put the findings on conversational turn counts in a broader context, we simulate equation (3) in the following way. We eliminate variability in conversational turns due to recording duration for all individuals and fix it (before standardization) to twelve hours. We use mean values for the estimation sample for all the other variables. We predict error terms $\hat{\epsilon}_{i,j,1}$ and store the estimated values for the coefficients $\hat{\beta}_j$. Then, we predict, for each individual, two values:

$$\hat{Y}^0_{i,CT,1} = \hat{\beta}_0 + \hat{\beta}_2 \bar{Y}_{CT,0} + \bar{X}\hat{\beta}_3 + \hat{\beta}_4 \bar{R}_1 + \hat{\beta}_5 \bar{R}_0 + \hat{\epsilon}_{i,j,1}$$

$$\hat{Y}^1_{i,CT,1} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 \bar{Y}_{CT,0} + \bar{X}\hat{\beta}_3 + \hat{\beta}_3 \bar{R}_1 + \hat{\beta}_4 \bar{R}_0 + \hat{\epsilon}_{i,j,1}$$

The variables $\hat{Y}^0_{i,CT,1}$ and $\hat{Y}^1_{i,CT,1}$ represent the predicted number of conversational turns without the LENA Start and with the LENA Start Program, respectively. We find that the distribution of $\hat{Y}^0_{i,CT,1}$ has a mean of around 295 conversational turns in a twelve-hour window, while the distribution of $\hat{Y}^1_{i,CT,1}$ has a mean of about 560 conversational turns in twelve hours. Therefore, one way to quantify the impact of the LENA Start Program is the difference between these two means, which means that the program adds 265 conversational turns per day.

Another way to quantify the impact of the Program is to use the parameters of the normative distribution for children who are 34 months, which is the average age of the children at the date of the follow-up recording session. According to Gilkerson and Richards (2008), the mean is 496, and the standard deviation is 313. These figures imply that the conversational turns would be 64% of a standard deviation below the normative mean without the LENA Start Program. With the LENA Start Program, conversational turns are 20% of a standard deviation *above* the normative mean. Alternately, these values imply that families move from the 26th to the 58th percentile. These are enormous impacts on conversational turns.

These simulations suggest significant effects of the LENA Start Program on conversational turns if we control for conversational turns at baseline. To address these pre-existing differences, we then estimate a lagged dependent model. This specification is defensible if the heterogeneity is persistent and evolves in non-parallel trends (so the differences vary over time). Another specification is the fixed-effect (FE) model, which allows for persistent parallel trends in conversational turns (in the absence of the program). To verify the robustness of our findings, let $t \in \{0,1\}$ indicate the baseline ($t = 0$) and endline ($t = 1$) waves, and let $Z_{i,t} = t \times Z_i$. Consider the following specification:

$$Y_{i,t} = \beta_0 + \beta_1 Z_{i,t} + \beta_2 t + \beta_3 X_{i,t} + \beta_4 R_{i,t} + \eta_i + \epsilon_{i,t}. \quad (9)$$

In specification (9), we absorb the time-invariant characteristics (such as demographic characteristics and randomization groups) into the individual fixed effect term η_i. The only demographic characteristic that varies over time is the child's age, which we include in $X_{i,t}$ in (9). As Table F1 shows, the fixed-effect estimator of the impact of the LENA Start Program is 32.3% of a standard deviation, and it is statistically significant at the 10% level.

Next, we expand on the analysis above by estimating an instrumental variable fixed effect (IV-FE) model. Let $D_{i,t} = t \times D_i$. Moreover, consider the following specification:

$$D_{i,t} = \alpha_0 + \alpha_1 Z_{i,t} + \alpha_2 t + \alpha_3 X_{i,t} + \alpha_4 R_{i,t} + \eta_i + \omega_{i,1} \quad (10)$$

$$Y_{i,t} = \beta_0 + \beta_1 D_{i,t} + \beta_2 t + \beta_3 X_{i,t} + \beta_4 R_{i,t} + \eta_i + \epsilon_{i,t}. \quad (11)$$

We present the results of the IV-FE model in Table F1. As estimated by the IV-FE procedure, the program's impact is 57.8% of a standard deviation and is statistically significant at the 10% level. Therefore, once we account for the differences in conversational turns at baseline, we find that the LENA Start Program impacts conversational turns between parents and children, and the magnitudes of the impacts are substantial.

Finally, we also conduct additional exercises to investigate the robustness of our point estimates. First, we investigate the sensitivity of the ITT estimates concerning sample attrition due to invalid recording data. We address sample attrition in multiple ways. First, we use inverse probability weighting (IPW), which assumes that attrition is exogenous. Second, we estimate a Heckman selection model, thus allowing for endogenous attrition. Third, we estimate the sharp bounds Lee (2009) proposed, which also allow for endogenous attrition. Our results are not sensitive to attrition. The IPW procedure produces a slightly lower point estimate for the ITT (0.295) and a t-statistic of 1.95. The Heckman selection procedure generates an ITT estimate of 0.314, which is statistically significant at 5%. The lower and upper bounds of ITT are 0.09 and 0.33, respectively. It is impossible to reject the null (no impact) for the lower bound, but the confidence interval for the upper bound does not include zero. If we include the whole sample (i.e., the 289 eligible families), ITT's lower and upper bounds are 0.06 and 0.37, respectively. We again detect the same significance level pattern for that exercise. See Table F2.

Second, we also investigate how the ITT and LATE change once we control for recording duration in flexible ways. As we show in Table F3, the point estimates are robust to the way we control for recording duration. If anything, our estimates of the program impacts and their standard errors decrease as we specify recording duration in more flexible ways.

G. Additional Results on Adult Words and Exposure to TV

Now, we discuss the impact of the LENA Start Program on adult words. The LENA System produces several counts of adult words. First, it automatically generates an overall adult word count. This automatic count combines words addressed to the child and speech overheard by the child. Because the literature emphasizes the role of speech addressed to the child in

events of joint attention or speech recasting, we use the ADEX software to isolate adult words in audio segments that have conversations between the focus child and a female or male adult. By imposing this additional constraint, we increase the likelihood that the adult words are part of child-directed speech. As we have done so far, we use the lagged dependent variable models described in equations (1) and (2)-(3), but the results are the same if we use the FE or IV-FE procedures described in equations (9) and (10)-(11).

Table G1 presents the results. We divide the table into two panels, and, within each panel, we present both the results for the ITT (OLS Estimator) and LATE (2SLS Estimator). Panel A displays the number of adult words in audio segments with the focus child and a female or male adult. The OLS results are positive (not statistically significant) but moderate in size. When we focus on the 2SLS results, we see that the LENA Start Program has a large and significant impact on adult words. When we parse out words spoken by female adults and the ones spoken by male adults, we find that male adults, not female adults, drive the results. This finding is somewhat surprising because only one of the LENA Start Program attendants was a male, and all the others were female. This finding suggests spillover of the program to the male adult in the household.

Panel B in Table G1 focuses on the audio segments initiated by the child and compares the adult words in those segments between the control and invitation (OLS) or LENA Start attendance group (2SLS) groups. Both estimators suggest large differences in adult words: once a child initiates a conversation, the adults in the attendance group have responses with more words spoken to the child. The latter finding justifies returning to conversational turns in segments with the focus child and an adult person

In sum, the LENA Start Program increases conversational turns and adult words spoken to the child. The more significant number of conversational turns and words in turns is the product of a higher number of conversational turns initiated by the child and responded to by parents in ways that do not stop the conversation but rather allow the child to continue to talk and interact.

Next, we also investigated the impact of the LENA Start Program on the focus child's exposure to audio from the TV or other electronics. Table G2 shows the results. As in other tables in the paper, Panel A reports the ITT estimates, while Panel B documents the LATE estimates. We do not find any impact of the program on exposure to TV. The point estimates in Model 1 indicate an increase. The sign of the estimates turns negative (denoting a reduction in exposure to TV) once we control for demographic characteristics (Model 2). If we also include

statistics about TV/Electronics at baseline, the coefficient becomes even smaller, but it is still statistically insignificant. We conclude that the LENA Start Program does not reduce exposure to TV/Electronics.

Table B1: Characteristics of PHD Study by Residence

	Inner-City Sample		Suburban Sample		p-value
	(1)		(2)		(1) vs (2)
VARIABLES	N	Mean	N	Mean	
Household income is less than or equal to twice the federal poverty line	652	0.750 (0.430)	165	0.050 (0.230)	0.000
Mother has attended some postsecondary program	657	0.080 (0.270)	165	0.510 (0.500)	0.000
Mother is Hispanic	657	0.150 (0.360)	165	0.040 (0.190)	0.000
Mother is Non-Hispanic black	657	0.640 (0.480)	165	0.150 (0.350)	0.000
Mother is Non-Hispanic white	657	0.140 (0.350)	165	0.780 (0.420)	0.000
Mother is single	657	0.720 (0.450)	165	0.170 (0.380)	0.000
Mother is cohabiting	657	0.110 (0.310)	165	0.040 (0.200)	0.001
Mother is married	657	0.180 (0.380)	165	0.790 (0.410)	0.000
Standardized HOME Score at 9 months	547	-0.170 (1.050)	140	0.660 (0.290)	0.000
Conversational Turns at 9 months	159	306.217 (150.268)	80	343.958 (168.755)	0.092
Adult Word Counts at 9 months	159	15260.390 (7367.602)	80	14559.220 (7203.537)	0.481
Seconds of Exposure to TV	159	6705.031 (4693.314)	80	3084.132 (2551.884)	0.000
Standardized Language Score from the Bayley Scale	541	-0.561 (0.740)	133	0.213 (0.818)	0.000

Notes: Standard deviation in parenthesis. The p-values are obtained from item-wise regressions using the variables labeled in the first column as the dependent variables and the urban area indicator as the only independent variable. These are all robust uncertainty estimates.

Table B2: Between and Within Sum of Squares as Fractions of Total Sum of Squares (PHD Study)

VARIABLES	Between	Within
Standardized HOME Score	11.20%	88.80%
Conversational Turn Counts (12 hours)	1.30%	98.70%
Standardized BSID Language Composite Score	14.30%	85.70%

Notes: This table shows estimates of between and within variation in investment and child development measures for the PHD Study sample.

Table B3: Correlation between HOME and Conversational Turns with Quartiles of Family Income (PHD Study)

VARIABLES	(1) HOME Score	(2) Conversational Turns
Second quartile of family income	0.591***	0.095
	(0.170)	(0.164)
Third quartile of family income	0.933***	0.412**
	(0.164)	(0.191)
Fourth quartile of family income	1.068***	0.011
	(0.147)	(0.155)
Observations	234	239
R-squared	0.196	0.094

Notes: Robust standard errors in parentheses. In the regression of conversational turns against quartiles of family income, we also control the recording session's duration. *** p<0.01, ** p<0.05, * p<0.1.

Table B4: Correlation between Language Development with HOME and LENA Measures (PHD Study)

	Dependent Variable: Standardized Bayley Scales of Infant Development Language Composite Score			
VARIABLES	Model 1	Model 2	Model 3	Model 4
	Panel A: Standardize HOME Score Only			
Standardized HOME Score	0.261***	0.099**	0.084**	0.070*
	(0.065)	(0.044)	(0.042)	(0.042)
Observations	223	223	223	223
	Panel B: Standardized Conversational Turns Only			
Standardized Conversational Turn Counts	0.193***	0.165***	0.153***	0.142***
	(0.066)	(0.058)	(0.057)	(0.055)
Observations	223	223	223	223
	Panel C: Standardized Adult Word Counts Only			
Standardized Adult Word Counts	0.100*	0.102*	0.101*	0.088
	(0.059)	(0.054)	(0.054)	(0.055)
Observations	223	223	223	223
	Panel D: Standardized TV Time			
Standardized TV Time	-0.170***	-0.013	-0.002	-0.004
	(0.059)	(0.057)	(0.057)	(0.057)
Observations	223	223	223	223
Demographic characteristics	N	Y	Y	Y
Dummy for Inner-City Sample	N	N	Y	Y
Dummies for Quartiles of Family Income	N	N	N	Y
	Panel E: Combining Multiple Measures			
Standardized HOME Score	0.052	0.058	0.072*	0.053
	(0.041)	(0.042)	(0.044)	(0.043)
Standardized Conversational Turn Counts	0.137**			0.148*
	(0.055)			-0.077
Standardized Adult Word Counts		0.081		(0.017)
		(0.056)		(0.076)
Standardized TV Time			-0.012	-0.002
			(0.058)	(0.060)
Observations	223	223	223	223
Demographic characteristics	Y	Y	Y	Y
Dummy for Inner-City Sample	Y	Y	Y	Y
Dummies for Quartiles of Family Income	Y	Y	Y	Y

Notes: Robust standard errors in parentheses. Demographic information include a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording. *** p<0.01, ** p<0.05, * p<0.1.

Table C1: Family Location, Family Consenting, and Family Attendance to LENA Start Program

	Invitation to LENA Start (1)		Control (2)		p-value (1) vs (2)	
VARIABLES	N	Mean	N	Mean	Robust	Clustered
Family was located conditional on being eligible	144	0.660 (0.475)	145	0.648 (0.479)	0.839	0.812
Family consented conditional on being eligible and located	92	0.707 (0.458)	87	0.816 (0.390)	0.086	0.074
Family consented conditional on being eligible	144	0.451 (0.499)	145	0.490 (0.502)	0.516	0.470
Family attended LENA Start Program conditional on being eligible and having consented	65	0.538 (0.502)	71	0.056 (0.232)	0.000	0.001
Family attended LENA Start Program conditional on being eligible	144	0.243 (0.430)	145	0.028 (0.164)	0.000	0.000

Notes: Standard deviation in parenthesis. The p-values are obtained from item-wise regressions using the variables labeled in the first column as the dependent variables and the random assignment indicator as the only independent variable. We report the p-values for both robust and clustered uncertainty estimates clustered at the recruiment group level.

Table C2: First-stage Results of the LATE Strategy

	Dependent Variable: Attendance to the LENA Start Program		
Variables	Model 1	Model 2	Model 3
Random assignment to control or invitation arm	0.560***	0.561***	0.569***
Clustered SE	(0.123)	(0.123)	(0.107)
Robust SE	(0.074)	(0.080)	(0.092)
Observations	104	102	90
Kleibergen-Paap Wald rk F-statistic	57.970	49.120	38.260
R-squared	0.387	0.491	0.553
Recording duration at endline	Y	Y	Y
Dummies for Saturday and Sunday at endline	Y	Y	Y
Demographic characteristics	N	Y	Y
Conversational Turn Counts at baseline	N	N	Y
Recording duration at baseline	N	N	Y
Dummies for Saturday and Sunday at baseline	N	N	Y

Notes: Standard errors in parentheses. We report clustered and robust standard errors, and we report the most conservative significance level. Clustered standard errors are clustered at the recruiment group level. We add the following variables to control for differences in recording sessions. First, we control for recording duration in endline, a dummy variable for a recording done on Saturday, and another dummy variable for a recording done on Sunday (*Model 1)* . Second, we additionally add demographic characteristics (*Model 2)* including a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording; and dummy for randomization group (block). Third, we additionally add conversational turn counts at baseline, recoding duration at baseline, and dummies for Saturday and Sunday at baseline (*Model 3).*

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table D1: Predictors of Provision of Valid Recording Data and Total Duration of Recording Data

	Baseline			Endline		
	Probit	OLS	Heckman selection model	Probit	OLS	Heckman selection model
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Dummy for Valid Recording	Standardized Recording Duration	Standardized Recording Duration	Dummy for Valid Recording	Standardized Recording Duration	Standardized Recording Duration
Random assignment to control or invitation arm	0.215	-0.029	-0.040	0.049	0.405	0.406**
Clustered SE	(0.445)	(0.201)	(0.164)	(0.311)	(0.266)	(0.181)
Robust SE	(0.314)	(0.179)		(0.292)	(0.189)	
Participation in the Language Study at age 9 months	Y	N	N	Y	N	N
Observations	112	108	134	116	102	128
R-squared		0.322			0.366	

Notes: Standard errors in parentheses. We report both robust and clustered standard errors, except for the Heckman selection model. We also report the most conservative significance level. We add the following control variables to all of the models: a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording; and dummy for randomization group (block). In Columns (1) and (3), we additionally control for participation status in the Language Study at age 9 months. *** p<0.01, ** p<0.05, * p<0.1.

Table D2: Conversational Turn Counts at Baseline

	Model 1	Model 2	Model 3
VARIABLES	Panel A: OLS Estimator		
Random assignment to control or invitation arm	-0.199	-0.116	-0.123
Clustered SE	(0.170)	(0.179)	(0.099)
Robust SE	(0.190)	(0.159)	(0.133)
Observations	110	108	108
R-squared	0.102	0.390	0.526
	Panel B: 2SLS Estimator		
Attendance to the LENA Start program	-0.397	-0.238	-0.249
Clustered SE	(0.329)	(0.288)	(0.146)
Robust SE	(0.360)	(0.290)	(0.237)
Observations	110	108	108
R-squared	0.108	0.409	0.531
Dummies for Randomization Group	Y	Y	Y
Demographic variables	N	Y	Y
Recording duration	N	N	Y
Dummies for Saturday/Sunday	N	N	Y

Notes: Standard errors in parentheses. We report both robust and clustered standard errors, and we report the most conservative significance level. We add the following demographic variables to Models 2-3: a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording. We add dummies for randomization group (block) in all Models 1-3. We add recording duration and dummies for recordings that took place on a Saturday or a Sunday in Model 3. *** p<0.01, ** p<0.05, * p<0.1.

Table E1: Sensitivity of Impacts on Maternal Beliefs to Extreme Percentiles

VARIABLES	5th and 95th percentiles	10th and 90th percentiles
	Panel A: OLS Estimator	
Random assignment to control or invitation arm	0.318*	0.265*
Clustered SE	(0.157)	(0.134)
Robust SE	(0.160)	(0.127)
Observations	128	128
	Panel B: 2SLS Estimator	
Attendance to the LENA Start program	0.608**	0.505**
Clustered SE	(0.256)	(0.215)
Robust SE	(0.286)	(0.226)
Observations	128	128

Notes: Standard errors in parentheses. We report both robust and clustered standard errors, and we report the most conservative significance level. We add the following variables to control for differences in survey responses. We add the lagged dependent variable. We also add a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording; and dummy for randomization group (block). *** p<0.01, ** p<0.05, * p<0.1.

Table E2: Results of Item Response Theory Analysis

	Panel A: Self Efficacy Scale				Panel B: Social Support Scale			
Item Discrimination Parameter	0.835*** (0.088)				0.599*** (0.073)			
	Item and Alternative Difficulty Parameter				Item and Alternative Difficulty Parameter			
	Item 1	Item 2	Item 3	Item 4	Item 1	Item 2	Item 3	Item 4
Alternative 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Alternative 2	1.301*		0.980**	1.405***	1.837***	0.969**	-0.248	0.159
	(0.737)		(0.395)	(0.422)	(0.682)	(0.402)	(0.542)	(0.401)
Alternative 3	3.808***	2.540***	2.879***	3.196***	3.694***	1.842***	1.931***	1.669***
	(0.657)	(0.522)	(0.390)	(0.430)	(0.651)	(0.417)	(0.409)	(0.362)
Alternative 4	5.595***	3.930***	3.435***	3.778***	4.732***	2.607***	2.736***	1.939***
	(0.687)	(0.546)	(0.430)	(0.471)	(0.665)	(0.434)	(0.431)	(0.398)
Alternative 5	6.561***	5.263***	3.525***	4.197***	5.272***	4.009***	4.023***	3.502***
	(0.705)	(0.558)	(0.440)	(0.481)	(0.669)	(0.428)	(0.431)	(0.385)
Variance of latent factor	1.000 (0.000)				1.000 (0.000)			
Observations	266	266	266	266	266	266	266	266

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table E3: Relationship Between Baseline Data for Mechanisms and Treatment Assignment

VARIABLES	Parental Self Efficacy	Parental Sense of Social Support	Parental Knowledge	Parental Beliefs	
				Error-Ridden Measure	Factor Score
Random assignment to control or invitation arm	0.198	0.048	0.188	0.281	0.217
Clustered SE	(0.157)	(0.191)	(0.115)	(0.199)	(0.213)
Robust SE	(0.181)	(0.190)	(0.145)	(0.193)	(0.188)
Observations	134	134	134	134	134
R-squared	0.211	0.114	0.457	0.100	0.237

Notes: Standard errors in parentheses. We report both robust and clustered standard errors, and we report the most conservative significance level. In all regressions, we control for demographic information including a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording; and dummy for randomization group (block). *** p<0.01, ** p<0.05, * p<0.1.

Table E4: Impact of the LENA Start Program by Subscale of the Parental Knowledge Questionnaire

VARIABLES	Learning to Talk	Reading Books	Learning Math	School Readiness	Language and TV
	Panel A: OLS Estimator				
Random assignment to control or invitation arm	0.117	0.223	0.359	0.250	0.189
Clustered SE	(0.183)	(0.112)	(0.215)	(0.156)	(0.171)
Robust SE	(0.168)	(0.166)	(0.155)	(0.149)	(0.163)
Observations	128	128	128	128	128
	Panel B: 2SLS Estimator				
Attendance to the LENA Start program	0.223	0.428	0.685*	0.478*	0.362
Clustered SE	(0.282)	(0.142)	(0.371)	(0.256)	(0.261)
Robust SE	(0.288)	(0.290)	(0.257)	(0.270)	(0.283)
Observations	128	128	128	128	128

Notes: Standard errors in parentheses. We report both robust and clustered standard errors, and we report the most conservative significance level. We add the following variables to control for differences in survey responses. We add the lagged dependent variable. We also add a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording; and dummy for randomization group (block). *** p<0.01, ** p<0.05, * p<0.1.

Table F1: The Impact of the LENA Start Program on Conversational Turns at Endline (Fixed Effects Model)

	Fixed Effect Estimator	IV Fixed Effect Estimator
VARIABLES	(1)	(2)
Impact of LENA Start (ITT and LATE)	0.323*	0.578*
Clustered SE	(0.179)	(0.152)
Robust SE	(0.179)	(0.301)
Observations	214	214
Number of participants	122	122

Notes: Standard errors in parentheses. We report both robust and clustered standard errors, and we report the most conservative significance level. In Column (1), we cluster standard errors at individual level. In Column (2), we cluster standard errors at the group level. We add the following variables in these two regression models. Child's age at the time of the recording, a dummy for a recording that took place on a Saturday, and a dummy for a recording that took place on a Sunday. *** p<0.01, ** p<0.05, * p<0.1.

Table F2: The Sensitivity of the Estimates of Intent-to-Treat Treatment Effect Parameters with Respect to Sample Attrition Due to Invalid Recording

VARIABLES	Inverse Probability Weight	Heckman Selection	Lee Sharp Bounds	Lee Sharp Bounds (whole sample)
Random assignment to LENA Start Program	0.295	0.314**		
Clustered SE	(0.150)	(0.149)		
Robust SE	(0.196)			
Lower Bound (Lee Sharp Bounds)			0.091	0.056
			(0.148)	(0.160)
Upper Bound (Lee Sharp Bounds)			0.330**	0.366**
			(0.152)	(0.158)
Observations	78	116	136	289

Notes: Standard errors in parentheses. We report both clustered and robust standard errors for Inverse Probability Weight model, and we report the most conservative significance level. We report clustered standard errors for the Heckman Selection model and the Lee Sharp Bounds. We use the following variables to predict attrition status: a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording; and dummy for randomization group (block). We add the same set of variables as controls to all of the models. *** p<0.01, ** p<0.05, * p<0.1.

Table F3: Sensitivity of Estimates of the Program Impact with Respect to Recording Duration Specification

	Fixed Effect Specification			
VARIABLES	(1) Linear	(2) Quadratic	(3) Cubic	(4) Quartic
	Panel A: OLS Estimator			
Random assignment to control or invitation arm	0.323*	0.326*	0.338*	0.323*
Clustered SE	(0.072)	(0.063)	(0.063)	(0.075)
Robust SE	(0.179)	(0.176)	(0.177)	(0.174)
Observations	214	214	214	214
Number of participants	122	122	122	122
	Panel B: 2SLS Estimator			
Attendance to the LENA Start program	0.578*	0.582**	0.600**	0.580*
Clustered SE	(0.152)	(0.142)	(0.149)	(0.167)
Robust SE	(0.301)	(0.297)	(0.298)	(0.299)
Observations	214	214	214	214
Number of participants	122	122	122	122

Notes: Standard errors in parentheses. We report both clustered and robust standard errors, and we report the most conservative significance level. We add the following variables in all regressions: Child's age at the time of the recording, a dummy for a recording that took place on a Saturday, and a dummy for a recording that took place on a Sunday. *** p<0.01, ** p<0.05, * p<0.1.

Table G1

Impact of the LENA Start Program on Adult Words in Audio Segments with Key Child and an Adult Person

	Panel A: All Audio Segments with Focus Child and an Adult Person		
	Adult Words	Female Adult Words	Male Adult Words
VARIABLES		OLS Estimator	
Random assignment to control or invitation arm	0.263	0.223	0.336
Clustered SE	(0.169)	(0.203)	(0.217)
Robust SE	(0.163)	(0.171)	(0.178)
Randomization inference p-value	0.175	0.249	0.098
Observations	87	87	87
		2SLS Estimator	
Attendance to the LENA Start program	0.468**	0.392	0.587**
Clustered SE	(0.200)	(0.260)	(0.266)
Robust SE	(0.230)	(0.240)	(0.251)
Observations	87	87	87
	Panel B: Audio Segments Initiated by the Child and Followed by Adult Person		
	Adult Words	Female Adult Words	Male Adult Words
		OLS Estimator	
Random assignment to control or invitation arm	0.495**	0.487*	0.388**
Clustered SE	(0.180)	(0.220)	(0.156)
Robust SE	(0.181)	(0.182)	(0.185)
Randomization inference p-value	0.007	0.008	0.049
Observations	87	87	87
		2SLS Estimator	
Attendance to the LENA Start program	0.886***	0.850***	0.684**
Clustered SE	(0.220)	(0.269)	(0.184)
Robust SE	(0.272)	(0.266)	(0.270)
Observations	87	87	87

Notes: Standard errors in parentheses. We report both robust and clustered standard errors, and we report the most conservative significance level. We also report the randomization inference p-value after 2,000 stratified clustered resampling. We add the following variables to control for differences in recording sessions. First, we add the lagged dependent variable and we control for recording duration in baseline and endline. Second, we add a dummy variable for a recording done on Saturday and another dummy variable for a recording done on Sunday. Additionally, we add a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording; and dummy for randomization group (block). *** p<0.01, ** p<0.05, * p<0.1.

Table G2: Impact of the LENA Start Program on Exposure to TV/Electronics

VARIABLES	Model 1	Model 2	Model 3
	Panel A: OLS Estimator		
Random assignment to control or invitation arm	0.046	-0.050	-0.128
Clustered SE	(0.168)	(0.210)	(0.212)
Robust SE	(0.202)	(0.229)	(0.245)
Randomization inference p-value	0.823	0.813	0.582
Observations	104	102	90
R-squared	0.196	0.318	0.451
	Panel B: 2SLS Estimator		
Attendance to the LENA Start program	0.082	-0.089	-0.216
Clustered SE	(0.269)	(0.309)	(0.272)
Robust SE	(0.335)	(0.361)	(0.350)
Observations	104	102	90
R-squared	0.199	0.313	0.444

Notes: Standard errors in parentheses. We report clustered and robust standard errors, and we report the most conservative significance level. We also report the randomization inference p-value after 2,000 stratified clustered resampling. We add the following variables to control for differences in recording sessions. First, we control for recording duration in endline, a dummy variable for a recording done on Saturday, and another dummy variable for a recording done on Sunday (Model 1). Second, we additionally add demographic characteristics (Model 2) including a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording; and dummy for randomization group (block). Third, we additionally add conversational turn counts at baseline, recoding duration at baseline, and dummies for Saturday and Sunday at baseline (Model 3). *** p<0.01, ** p<0.05, * p<0.1.

TABLE G3
Mediation Analysis Results

	Dependent Variable: Conversational Turn Counts					
		Mediators				
	ITT	+ Parental Sense of Social Support	+ Parental Self-Efficacy	+ Parental Knowledge	+ Parental Beliefs Error-Ridden	+ Parental Beliefs Factor Score
	(1)	(2)	(3)	(4)	(5)	(6)
Random assignment to control or invitation arm	0.314**	0.316**	0.306**	0.212	0.253*	0.163
Clustered SE	(0.113)	(0.126)	(0.132)	(0.124)	(0.137)	(0.147)
Robust SE	(0.152)	(0.152)	(0.151)	(0.136)	(0.148)	(0.145)
Parental sense of social support		0.041				
Clustered SE		(0.041)				
Robust SE		(0.060)				
Parental self-efficacy			-0.041			
Clustered SE			(0.079)			
Robust SE			(0.082)			
Parental knowledge				0.198		
Clustered SE				(0.173)		
Robust SE				(0.132)		
Parental beliefs (error-riden)					0.098	
Clustered SE					(0.094)	
Robust SE					(0.085)	
Parental beliefs (factor score)						0.198**
Clustered SE						(0.074)
Robust SE						(0.099)
Observations	90	90	90	90	90	90
Control Group Mean	-0.089	-0.089	-0.089	-0.089	-0.089	-0.089
Monte Carlo 95% CIs for Indirect Effects (clustered)		[-0.053; 0.014]	[-0.080; 0.057]	[-0.060; 0.284]	[-0.043; 0.168]	[0.026; 0.288]
Monte Carlo 95% CIs for Indirect Effects (robust)		[-0.086; 0.035]	[-0.060; 0.038]	[-0.027; 0.262]	[-0.035; 0.169]	[0.002; 0.321]

NOTE. — Standard errors in parentheses. We report clustered and robust standard errors, and we report the most conservative significance level. Clustered standard errors are clustered at the recruiment group level. In all the regressions, we add the following variables to control for differences in survey responses. We add the lagged dependent variable. Additionally, we add a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording; and dummy for randomization group (block). The numbers in square brackets show the upper limits and lower limits of the Monte Carlo 95% confidence intervals for the indirect effects of the LENA Start program on the conversational turns through the respective mediators.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

References

Andrew, Alison, Orazio Attanasio, Emla Fitzsimons, Sally M. Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina. "Impacts 2 Years after a Scalable Early Childhood Development Intervention to Increase Psychosocial Stimulation in the Home: A Follow-up of a Cluster Randomised Controlled Trial in Colombia." *PLoS Medicine* 15, no. 4 (2018).

Attanasio, Orazio, Flavio Cunha, and Pamela Jervis. "Subjective Parental Beliefs: Their Measurement and Role." In *NBER Working Paper 26516*, 2019.

Attanasio, Orazio P., Camila Fernandes, Emla O. A. Fitzsimons, Sally M. Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina. "Using the Infrastructure of a Conditional Cash Transfer Program to Deliver a Scalable Integrated Early Child Development Program in Colombia: Cluster Randomized Controlled Trial." *British Medical Journal* 349, no. g5785 (2014).

Bailey, Drew, Greg Duncan, Flavio Cunha, Barbara Foorman, and David Yeager. "Persistence and Fadeout of Educational Intervention Effects: Mechanisms and Potential Solutions." *Psychological Science in the Public Interest* 21, no. 2 (2020): 55-97.

Bandura, A. "Self-Efficacy - toward a Unifying Theory of Behavioral Change." *Psychological Review* 84, no. 2 (1977): 191-215.

Bhalotra, Sonia, Adeline Delavande, Paulino Font Gilabert, and Joanna Maselko. "Maternal Investments in Children: The Role of Expected Effort and Returns." In *IZA Institute of Labor Economics Discussion Paper 13056*, 2020.

Biroli, Pietro, Teodora Boneva, Akash Raja, and Chris Rauh. "Parental Beliefs About Returns to Child Health Investments." *Journal of Econometrics* 231, no. 1 (2022): 33-57.

Boneva, Teodora, and Chris Rauh. "Parental Beliefs About Returns to Educational Investments—the Later the Better?" *Journal of the European Economic Association,* 16, no. 6 (2018): 1669-711.

Bradley, Robert H., and Bettye M. Caldwell. "The Relation of Home-Environment, Cognitive Competence, and Iq among Males and Females." *Child Development* 51, no. 4 (1980): 1140-48.

Bruner, Jerome. *Child's Talk: Learning to Use Language.* . New York, NY: W.W. Norton & Co., 1983.

Carneiro, Pedro, Emanuela Gallaso, Italo L. García, Paula Bedregal, and Miguel Cordero. "Parental Beliefs, Investments, and Child Development: Evidence from a Large-Scale

Experiment." In *Discussion Paper Series*, edited by IZA Institute of Labor Economics, 2019.

Cleave, Patricia L., Stephanie D. Becker, Maura K. Curran, Amanda J O Van Horn, and Marc E Fey. "The Efficacy of Recasts in Language Intervention: A Systematic Review and Meta-Analysis " *American Journal of Speech-Language Pathology* 24 (2015): 237-55.

Coleman, Priscilla K., and Katherine H. Karraker. "Self-Efficacy and Parenting Quality: Findings and Future Applications." *Developmental Review* 18, no. 1 (1998): 47-85.

Cunha, Flavio, Irma Elo, and Jennifer Culhane. "Eliciting Maternal Expectations About the Technology of Cognitive Skill Formation." In *NBER Working Paper 19144*, 2013.

———. "Maternal Subjective Expectations About the Technology of Skill Formation Predict Investments in Children One Year Later." *Journal of Econometrics* 231, no. 1 (2022): 3-32.

Duncan, Greg, and Richard J. Murnane. *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances (Copublished with the Spencer Foundation)* New York, NY.: Russell Sage Foundation, 2011.

Fernald, Anne, Virginia A. Marchman, and Adriana Weisleder. "Ses Differences in Language Processing Skill and Vocabulary Are Evident at 18 Months." *Developmental Science* 16, no. 2 (2012): 234-48.

Foorman, Barbara R. , Yaacov Petscher, and Sarah Herrera. "Unique and Common Effects of Decoding and Language Factors in Predicting Reading Comprehension in Grades 1–10." *Learning and Individual Differences* 63, no. 1 (2018): 12-23.

García, Jorge Luis, and James J. Heckman. "Parenting Promotes Social Mobility within and across Generations." In *NBER Working Paper*, 2022.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science* 344, no. 6187 (2014): 998-1001.

Gertler, Paul, James J Heckman, Rodrigo Pinto, Susan M Chang, Sally M Grantham-McGregor, Christel Vermeersch, Susan Walker, and Amika Wright. "Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31." In *NBER Working Paper*, 2021.

Gilkerson, Jill, and Jeffrey A. Richards. "The Lena Natural Language Study." 2008.

Gilkerson, Jill, Jeffrey A. Richards, Steven F. Warren, D. Kimbrough Oller, Rosemary Russo, and Betty Vohr. "Language Experience in the Second Year of Life and Language Outcomes in Late Childhood." *Pediatrics* 142, no. 4 (2018).

Golinkoff, Roberta M., Erika Hoff, Meredith L. Rowe, Catherine S. Tamis-LeMonda, and Kathy Hirsh-Pasek. "Language Matters: Denying the Existence of the 30-Million-Word Gap Has Serious Consequences." *Child Development* 0, no. 0 (2019).

Gough, Philip B., and William E. Tunmer. "Decoding, Reading, and Reading Disability." *Remedial and Special Education* 7, no. 1 (1986): 6-10.

Grantham-McGregor, Sally, Christine A. Powell, Susan P. Walker, and John H. Himes. "Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: The Jamaican Study." *The Lancet* 338, no. 8758 (1991): 1-5.

Hamadani, Jena D., Syed N. Huda, Fahmida Khatun, and Sally M. Grantham-McGregor. "Psychosocial Stimulation Improves the Development of Undernourished Children in Rural Bangladesh." *Journal of Nutrition* 136, no. 10 (2006): 2645-52.

Hanushek, Erik A., Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann. "Returns to Skills around the World: Evidence from Piaac." *European Economic Review* 73 (2015): 103-30.

Hart, Betty, and Todd Risley. *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H. Brookes Publishing Company, 1995.

Heckman, James J. , Margaret L. Holland, Kevin K. Makino, Rodrigo Pinto, and Maria Rosales-Rueda. "An Analysis of the Memphis Nurse-Family Partnership Program." In *NBER Working Paper 23610*, 2017.

Heckman, James J. , Bei Liu, Mai Lu, and Jin Zhou. "The Impacts of a Prototypical Home Visiting Program on Child Skills." In *NBER Working Paper 27356*, 2020.

Hoff, Erika. "How Social Contexts Support and Shape Language Development." *Developmental Review* 26, no. 1 (2006): 55-88.

———. "The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech." *Child Development* 74, no. 5 (2003): 1368-78.

Huttenlocher, Janellen, Marina Vasilyeva, Heidi R. Waterfall, Jack L. Vevea, and Larry V. Hedges. "The Varieties of Speech to Young Children." *Developmental Psychology* 43, no. 5 (2007): 1062-83.

Huttenlocher, Janellen, Heidi R. Waterfall, Marina Vasilyeva, Jack L. Vevea, and Larry V. Hedges. "Sources of Variability in Children's Language Growth." *Cognitive Psychology* 61, no. 4 (2010): 343-65.

Jones, Tracy L., and Ronald J. Prinz. "Potential Roles of Parental Self-Efficacy in Parent and Child Adjustment: A Review." *Clinical Psychology Review* 25, no. 3 (2005): 341-63.

Kalil, Ariel "Inequality Begins at Home: The Role of Parenting in the Diverging Destinies of Rich and Poor Children." In *Families in an Era of Increasing Inequality: Diverging Destinies*, edited by Paul R. Amato, Alan Booth, Susan M. McHale and Jennifer Van Hook, 63-82. New York: Springer, 2014.

Leahy-Warren, Patricia, Geraldine McCarthy, and Paul Corcoran. "First-Time Mothers: Social Support, Maternal Parental Self-Efficacy, and Postnatal Depression." *Journal of Clinical Nursing* 21, no. 3-4 (2012): 388-97.

Lee, David S. . "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76, no. 3 (2009): 1071-102.

Leech, Kathryn A., Ran Wei, Jeffrey R Harring, and Meredith L Rowe. "A Brief Parent-Focused Intervention to Improve Preschoolers' Conversational Skills and School Readiness." *Developmental Psychology* 54, no. 1 (2018): 15-28.

List, John A., Julie Pernaudet, and Dana L. Suskind. "It All Starts with Beliefs: Addressing the Roots of Educational Inequities by Changing Parental Beliefs." edited by NBER Working Paper, 2021.

McGillion, Michelle , Julian M Pine, Jane S Herbert, and Danielle Matthews. "A Randomised Controlled Trial to Test the Effect of Promoting Caregiver Contingent Talk on Language Development in Infants from Diverse Socioeconomic Status Backgrounds." *Journal of child psychology and psychiatry, and allied disciplines* 58, no. 10 (2017): 1122-31.

Messenger, K., S. Yuan, and C. Fisher. "Learning Verb Syntax Via Listening: New Evidence from 22-Month-Olds." *Language Learning and Development* 11, no. 4 (2015): 356-68.

Murnane, Richard J., John B. Willett, Yves Duhaldeborde, and John H. Tyler. "How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?" *Journal of Policy Analysis and Management* 19, no. 4 (2000): 547-68.

OECD. "Skills Matter: Further Results from the Survey of Adult Skills." In *OECD Skills Studies*. Paris, 2016.

Olds, David L, Harriet Kitzman, Robert Cole, JoAnn Robinson, Kimberly Sidora, Dennis W Luckey, Charles R Henderson Jr, Carole Hanks, Jessica Bondy, and John Holmberg. "Effects of Nurse Home-Visiting on Maternal Life Course and Child Development: Age 6 Follow-up Results of a Randomized Trial." *Pediatrics* 114, no. 6 (2004): 1550-59.

Pan, Barbara A., Meredith L. Rowe, Judith D. Singer, and Catherine E. Snow. "Maternal Correlates of Growth in Toddler Vocabulary Production in Low-Income Families." *Child Development* 76, no. 4 (2005): 763-82.

Ramirez, Naja Ferjan, Sarah Roseberry Lytle, and Patricia K. Kuhl. "Parent Coaching Increases Conversational Turns and Advances Infant Language Development." *Proceedings of the National Academy of Sciences of the United States of America* 117, no. 7 (2020): 3484-91.

Rollins, Pamela R. "Caregivers' Contingent Comments to 9-Month-Old Infants: Relationships with Later Language." *Applied Psycholinguistics* 24, no. 2 (2003): 221-34.

Romeo, R. R., J. A. Leonard, S. T. Robinson, M. R. West, A. P. Mackey, M. L. Rowe, and J. D. E. Gabrieli. "Beyond the 30-Million-Word Gap: Children's Conversational Exposure Is Associated with Language-Related Brain Function." *Psychological Science* 29, no. 5 (2018): 700-10.

Rowe, Meredith L. , and Kathryn A. Leech. "A Parent Intervention with a Growth Mindset Approach Improves Children's Early Gesture and Vocabulary Development." *Developmental science* 22, no. 4 (2019): e12792.

Rowe, Meredith L., and Susan Goldin-Meadow. "Early Gesture Selectively Predicts Later Language Learning." *Developmental Science* 12, no. 1 (2009): 182-87.

Scarborough, Hollis S, Susan Neuman, and David Dickinson. "Connecting Early Language and Literacy to Later Reading (Dis) Abilities: Evidence, Theory, and Practice." *Approaching difficulties in literacy development: Assessment, pedagogy and programmes* 10 (2009): 23-38.

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. "A Parent-Directed Language Intervention for Children of Low Socioeconomic Status: A Randomized Controlled Pilot Study." *Journal of Child Language* 43, no. 2 (2016): 366-406.

Suskind, Dana L., Kristin R. Leffel, Marc W. Hernandez, Shannon G. Sapolich, Elizabeth Suskind, Erin Kirkham, and Patric Meehan. "An Exploratory Study of “Quantitative Linguistic Feedback”: Effect of Lena Feedback on Adult Language Production." *Communication Disorders Quarterly* 34, no. 4 (2013): 199-209.

Sylvia, Sean , Nele Warrinnier, Renfu Luo, Ai Yue, Orazio Attanasio, Alexis Medina, and Scott Rozelle. "From Quantity to Quality: Delivering a Home-Based Parenting Intervention through China’s Family Planning Cadres." *The Economic Journal* 131, no. 635 (2021): 1365-400.

Tronick, Edward , Heidelise Als, Lauren Adamson, Susan. Wise, and T. Berry Brazelton. "The Infant's Response to Entrapment between Contradictory Messages in Face-to-Face Interaction." *Child and Adolescent Psychiatry* 17, no. 1 (1978): 1-13.

Vignoles, Anna, Augustin De Coulon, and Oscar Marcenaro-Gutierrez. "The Value of Basic Skills in the British Labour Market." *Oxford Economic Papers-New Series* 63, no. 1 (2011): 27-48.

Weisleder, A., and A. Fernald. "Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary." *Psychological Science* 24, no. 11 (2013): 2143-52.

Weitzman, Carol Cohen, Lisa Roy, Theodore Walls, and Ricarda Tomlin. "More Evidence for Reach out and Read: A Home-Based Study." *Pediatrics* 113, no. 5 (2004): 1248-53.

Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*: MIT press, 2010.

Young, Alwyn. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *The Quarterly Journal of Economics* 134, no. 2 (2018): 557-98.

Zhang, Yiwen, Xiaojuan Xu, Fan Jiang, Jill Gilkerson, Dongxin Xu, Jeffrey A. Richards, James Hamberger, and Keith Topping. "Effects of Quantitative Linguistic Feedback to Caregivers of Young Children: A Pilot Study in China." *Communication Disorders Quarterly* 37, no. 1 (2015): 16-24.