CESifo
AREA CONFERENCES 2020

Economics of Education

Online, 30 November–1 December 2020

Language Environment and Maternal Expectations: An Evaluation of the LENA Start Program

Flávio Cunha, Marsha Gerdes, and Snejana Nihtianova

CESifo

Language Environment and Maternal Expectations: An Evaluation of the LENA Start Program[*]

Flávio Cunha[†], Marsha Gerdes[‡], and Snejana Nihtianova[§]
October 17, 2020

Abstract

A large body of research documents that the children's early language environment predict their future language development and later literacy skills. In addition, research shows that parental beliefs about the returns to early investments predict investments in early human capital. In this paper, we report the results of an experimental evaluation of the LENA Start Program. We find that the program impacts the quality of the parental linguistic input. We provide evidence that indicates that parental beliefs is one of the mechanisms of the program.

Keywords: early childhood human capital formation, language development, parental beliefs.

---

[*] We thank Delitza Hernandez, Michelle Gifford, Shanae Smith, and Mary Triguba for excellent research assistance. We are thankful to all of the LENA Foundation team. Jill Gilkerson and Steve Hannon for all of the support and encouragement relating to the research aspects of this project. Traci Martin and Jess Simmons were excellent in supporting our implementation of the LENA Start Program at the Children's Hospital of Philadelphia. Their dedication to this program, which we have heard from several other sites in the nation, really energized our team. We are grateful to Jeffrey Richards who generously shared crucial knowledge on how to operationalize the data from the LENA System and provided valuable comments about this research. We thank participants at the following conferences: Sociedade Brasileira de Econometria in Natal, Brazil, CONBATRI in Juiz de Fora, Brazil, the Austin-Bergen Conference at UT Austin, and the briq Belief Conference at the University of Bonn. We thank seminar participants at the Federal Reserve Bank in New York City, EESP at FGV in Sao Paulo, Boston University, the University of Minnesota, CEPA Seminar at Stanford University, Vancouver School of Economics, and Rochester University This research was supported by grant 1R01HD073221-01A1 by the National Institute of Health. All remaining errors are ours.

[†] Rice University, Texas Policy Lab, and NBER.

[‡] Children's Hospital of Philadelphia.

[§] Rice University, Texas Policy Lab.

1. Introduction

The 2017 National Assessment of Educational Progress (NAEP, 2017) reports a sizeable socioeconomic gap in reading skills. In the United States, 46% of low-income[1] children enrolled in fourth grade have reading skills below the basic level. In contrast, the same rate for children raised in the middle- and high-income households is about 18%. Reading is not just critical for school success. Research shows that the capacity to read proficiently confers positive economic returns in the labor market. Hanushek, Schwerdt, Wiederhold, and Woessmann (2015) use the Programme for the International Assessment of Adult Competencies (PIAAC) data to estimate the return to numeracy, literacy, and problem-solving skills. They find that one standard deviation improvement in literacy skills raises hourly wage rates by approximately 7% even after controlling for potential experience, educational attainment, problem-solving skills, and numeracy skills. Their findings are consistent with those by Murnane, Willett, Duhaldeborde, and Tyler (2000) and Vignoles, De Coulon, and Marcenaro-Gutierrez (2011), who explore rich longitudinal data and document the same facts for the US and the UK, respectively. Beyond pecuniary returns, research shows that the ability to comprehend written text is associated with better access to health services,[2] improvements in self-esteem,[3] higher levels of empowerment[4] and political participation,[5] promotion of cultural change, and cultural diversity preservation.[6] Reading proficiency is so fundamental in modern life that it is considered a basic human right (UNESCO, 1975).

Gough and Tunmer (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.

[1] By low-income children we mean children whose family income is low enough that they are eligible for the National School Lunch Program.
[2] See, e.g., Sandiford et al. (1995) and Dexter, LeVine and Velasco (1998).
[3] Bown (1990) and Canieso-Doronila (1996).
[4] Burchfield et al. (2002) and Kagitcibasi, Goksen and Gulgoz (2005).
[5] Hannum and Buchmann (2003), Ireland (1994), Egbo (2000), and Purcell-Gates and Waterman (2000).
[6] Carr-Hill et al. (2001), Robinson-Pant (2000), Chebanne and Nyati-Ramahobo (2003), Norwood (2003), and Tarawa (2003).

Decoding, or word recognition, is the capacity to recognize print and read words written on a page. Language comprehension is the capacity to make sense of the language one hears or one reads. As Scarborough (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, Petscher, and Herrera (2018) provide compelling evidence about the importance of language proficiency in explaining variability in reading proficiency. The authors quantified language proficiency with listening comprehension measures, [7] vocabulary measures,[8] and syntax measures.[9] The team assessed decoding skills with assessments of phonological awareness[10] and tests of decoding fluency.[11] According to their findings, 80% of reading skills variability at Grade 10 is due to variability in language comprehension skills.[12] 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 short logical sentence comprehension. In sum, the socioeconomic gaps in language skills substantially predict the socioeconomic gaps in reading proficiency.

---

[7] 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).
[8] The team measured expressive vocabulary with FAIR Vocabulary Task (Grades 1 and 2), Peabody Picture Vocabulary Test-4 (PPVT-4, Dunn & Dunn, 2007, Grades 1-10) and the Study Aid and Reading Assistant (SARA, Sabatini, Bruce, and Steinberg, 2013).
[9] The authors measured receptive and expressive syntax with CELF-4.
[10] 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).
[11] The Test of Word Reading Efficiency–2 (TOWRE–2; Torgesen, Wagner, & Rashotte, 2012) is the instrument the team used to measure decoding fluency.
[12] 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.

A large body of research has documented large socioeconomic gradients in language development observed before school years. The lower the family's socio-economic background, the lower the child's language skill development, regardless of whether one considers language processing, language comprehension, or language production.[13] Children from high SES families score higher on standardized language tests that measure grammatical development and phonological awareness.[14] The lower the children's SES, the lower the sophistication of their narratives (Vernon-Feagans, Hammer, Miccio, & Manlove, 2001).

Research also suggests that early language environment influences early language comprehension skills and language development more broadly. Weisleder and Fernald (2013) show that the quality of a child's language environment[15] at baseline (age 19 months) predicts vocabulary and language processing skills[16] five months later, even after controlling for vocabulary and language processing skills at baseline. These findings are consistent with those reported by Pan, Rowe, Singer, and Snow (2005), who showed that the more complex the maternal linguistic input, the larger the child's vocabulary.[17] Gilkerson et al. (2018) report the results of a longitudinal study that first measures the quality of infants' and toddlers' language environment[18] and then, when the same children are nine to fourteen years old, measures their IQ and language development.[19] The authors

[13] Hart and Risley (1995), Huttenlocher et al. (2010), Fernald, Marchman and Weisleder (2012), Hoff (2006).
[14] Dollaghan et al. (1999), Morisset et al. (1990), McDowell, Lonigan and Goldstein (2007).
[15] 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.
[16] To measure language processing skills, the authors use the Looking-While-Listening task (LWL, Fernald et al., 2008). 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. Language processing skills is then measured as reaction time, that is, the amount of time the infant shifts away from the distracter to the target picture. Children with higher processing skills take less time to shift away to the target picture.
[17] See also Hoff (2003).
[18] 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.
[19] 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).

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[20], 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 vocabulary words with their children than low SES mothers. [21] Rowe and Goldin-Meadow (2009) found that high SES mothers gesture more about objects in the environment when they are close to their infant 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 lower SES mothers also tend to talk to direct their children's behavior, not to converse and 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.[22]

All in all, it is a desirable goal to have children become proficient readers, but a significant fraction of low-income children are failing to reach even basic levels of literacy because they do not have appropriate levels of language skills. The development of these skills requires greater exposure to language so that children can practice and hone their language processing skills and simultaneously increase their vocabulary. In this paper, we report the results of evaluating a scalable intervention designed to improve disadvantaged children's language environment. The intervention, known as the

---

[20] Broca's area.

[21] See also Huttenlocher, Vasilyeva, Waterfall, Vevea and Hedges (2007). All of these studies count words directed to the child and do not count words the child overhears in adult speech.

[22] Schieffelin and Ochs (1986), Hart and Risley (1995), Hoff (2006), Huttenlocher et al. (2007), Baker et al. (2001), Fernald, Marchman and Weisleder (2012), and Rowe (2012).

LENA Start Program[23], provides information about the importance of the language environment for language development, offers a menu of natural and practical actions that parents can take to improve their child's language environment, and provides objective feedback about the quality of the child's language environment for the duration of the intervention. These activities occur weekly for thirteen weeks. Unlike parenting programs, LENA Start is a group-centered model that reduces costs and may help parents build social capital and a strong social support network. [24, 25]

We report the results of an evaluation of the LENA Start Program that occurred within the Philadelphia Human Development (PHD) longitudinal study. The study recruited 822 mothers during the second trimester of their first pregnancy, and we elicited mothers' subjective beliefs about the importance of parent-child interaction for child development. When the children were nine to twelve months old, we visited the families in their homes and measured the quality of their home and language environment. When the children were two years old, we assessed several child development dimensions, including language development. For the LENA Start Program evaluation, we identified a group of 289 low-income families whose children's ages were appropriate for the intervention. We randomly allocated half of the families and children to the control group and the other half to the group who received an invitation to participate in the LENA Start Program. Then, we enrolled 136 families to participate in the LENA Start evaluation study.

The evaluation study's primary outcomes are objective measures of the language environment, which we obtained with the LENA System (see Gilkerson and Richards, 2017). This system allows researchers to measure several dimensions of the language environment: adult-child conversational turns, the number of adult words spoken around the child, and the amount of exposure to TV or other electronics. We find strong evidence that the LENA Start Program significantly impacts

[23] See a complete description of the program at https://www.lena.org/lena-start/.
[24] See also Providence Talks (2015), Zhang et al. (2015), and Gilkerson et al. (2017) for other interventions that use linguistic feedback to parents.
[25] As of Spring 2019, the LENA Start Program is being implemented by hospital systems, school districts, and NGOs that serve parents and their young children at 29 sites across North America. For a list of current LENA Start program sites, see https://www.lena.org/about/#where-is-lena .

conversational turns and adult words in conversations between the focus child and an adult. The Program does not impact exposure to TV or other electronics.

We collected data to study potential mechanisms that could explain any positive findings of the impact of the LENA Start Program on parent-child linguistic interaction. We find that the program’s impacts are explained by changing maternal beliefs about the language environment's importance for language development. We do not find evidence supporting other mechanisms we test with our data.

Our paper relates to the literature on parenting programs for language development in developmental psychology. Leech et al. (2018), McGillion et al. (2017), Suskind et al. (2013), Suskind et al. (2017), and Rowe and Leech (2019) have designed and evaluated parenting education programs that target malleable aspects of the parental language input. These interventions translate scientific information about how parental language input (a critical component of a child’s early language environment) predicts early language development. Also, these interventions employ parental coaching based on objective measures of the child’s language environment. As we discuss below, both ingredients are present in the LENA Start Program.

Our paper also contributes to the literature on parental beliefs. Cunha, Culhane, and Elo (2013) introduced an economic model in which the choice of how much to invest in the child’s early human capital depended on parental beliefs about the returns to investments. They elicited parental beliefs and found that low-income parents have low expectations about the returns to investments. Boneva and Rauh (2018) expanded this literature by showing that parental beliefs varied by the parents' socioeconomic status and showed heterogeneity in parental beliefs about the roles of early and late investments in children's human capital. Attanasio, Cunha, and Jervis (2019) and Cunha, Elo, and Culhane (2020) showed that parental beliefs predict investments in children's human capital. Finally, List, Pernaudet, and Suskind (2020) show that parental beliefs predict investments and that these beliefs are malleable. We find that beliefs are malleable and that changes in parental beliefs induced by participation in the program predict changes in parental language input choices.

We organize the rest of this paper in the following way. Section 2 presents previous findings from The Philadelphia Human Development Study (PHD) study and information about the LENA Start Program. Section 3 describes our study procedures. Section 4 presents results including the estimates of the impact of the LENA Start Program on statistics of the language environment and describes our findings of the potential mechanisms of the Program. Section 5 concludes.

2. The Philadelphia Human Development Study and LENA Start program

2.1 The Philadelphia Human Development Study

The Philadelphia Human Development (henceforth, PHD) Study is a longitudinal study aimed at measuring parental beliefs about the impact of early investments for early human capital formation. The study recruited 822 English-speaking women when they were 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 maternal beliefs and collected data about the study participants' demographic characteristics in the first visit. Table 1 presents the summary statistics of the two groups of PHD Study Participants (inner-city sample and suburban sample). The pregnant women enrolled in the Medicaid clinics are more likely to be poor, have lower educational attainment, are more likely to be Hispanic or Non-Hispanic black (and less likely to be Non-Hispanic white), and are more likely to be single or cohabiting (and less likely to be married) at the time enrollment.

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. In all of the 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 1, we show that the gap in the HOME score between the inner-city and suburban sample is about 83% of a standard deviation. Cunha, Elo, and Culhane (2020) combined data from the first and second rounds of the PHD study and found that maternal 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).

In the second visit, 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 have unique information about the quality of the early language environment that low-income children experience on a typical day when they were about one year old. Table 1 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 approximately 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 Hart and Risley (1995) show that the differences in the number of conversations initiated by children in high and low socioeconomic groups mostly 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 as well as speech that is not directed to the child but it is captured by the device because the adult is close to the child. Finally, the Inner-City sample children watched one extra hour of TV per day.

The third visit happened when the child was between twenty-two and twenty-six months old, at the Children's Hospital of Philadelphia research building. 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). We average these two scores to estimate a composite language score. We use the BSID III normalizing statistics to estimate the age-normed score. Table 1 shows that the children in the inner-city sample are 56% of a standard deviation below the mean in language development, while the children in the urban 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 the literature in the social sciences has extensively documented the differences between socioeconomic groups in investment and human capital have been extensively documented in the literature (e.g., see Duncan and Murnane, 2011). The data from the PHD Study showed that the within-group differences are even more considerable. Table 2 reports our estimates of between and within variation in investment and child development measures. For the HOME and the Bayley Language Composite Score, the within-group variation is 89% to 86% of the total variance, respectively. For conversational turns, the variation was almost entirely within groups. A possible result for 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 capture a mixture of permanent and transitory aspects of the child's language environment, and transitory shocks vary a lot within groups but little across groups. [26]

Another possibility is that the HOME scale captures the child environment's elements more strongly impacted by household income, while the LENA system is sensitive to parental behavior that is not influenced by family income. We regress standardized HOME, and standardized total conversational turn counts against family income quartiles to investigate this issue.[27] Indeed, as we present in Table 3, 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 in the second visit, with the language development measures we obtained in the third visit (about one year later). Our goal

[26] A challenge of using audio data obtained through one recording day is sampling variability. A child's language environment can fluctuate on a day-to-day basis a lot because of variation in adult's (or child's) moods, because of variation of which adults are in the household, and even because of the child's health (an ill child may sleep more hours, for example). To address sampling variability it would be desirable to record the audio environment for at least three or four days and then average across the days. This procedure, however, imposes burden on participant families and require appropriate funding because of the logistical costs of such an operation.

[27] We adjust for the duration of the recording when we include the total counts of conversational turns in the model.

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 4 shows the results. First, we estimate the correlation between our language development assessments with each language environment measure separately. For each of these one-to-one relationships, we estimate four models. The first model does not control for any observable characteristics of 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. The one-to-one relationships with language development are weaker, and not statistically significant once we control for observed characteristics.

Panel E combines the HOME with at least one LENA measure. All of the models in Panel E have controls for all of 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 we combine it with exposure to TV. In contrast, when we combine conversational turn counts with the HOME, the latter was not statistically significant, but the former is. Even when we consider 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, then the gap is more extensive (around 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 the inclusion of family characteristics or other measures of parent-child interactions. 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 4 are a lower bound for the language environment's contribution to predicting future language development. Below, when we describe the LENA Start Program, we summarize the literature that explains why conversational turns are so important for early 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. In the PHD Study data, the language environment's heterogeneity – measured by conversational turn counts – has a fragile association with family income. The study hypothesizes that parental linguistic input gaps are driven by heterogeneity in parental beliefs about the importance of the early language environment for a child's language development. In this paper, 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.

2.2 <u>The LENA Start Program</u>

The LENA Start Program aims to improve parental linguistic input by improving parental knowledge about the importance of early language environment for language development and providing tips for shaping an early language enviroment. The Program lasts thirteen weeks, and, during this period, groups of ten to twenty families meet for about one hour a week to build family engagement and social capital. The program has five components: Education, Coaching, Feedback, Book Reading, and Language Development Reporting.

<u>Education</u>: The program provides information about the importance of the language environment for a child's language development. As documented by Rowe (2008), parents

misperceive linguistic interaction with their young children as unimportant because young children do not yet know how to verbalize. The program aims to affect these beliefs by presenting parents with research that documents infants' capacity to engage in social interaction, and that parents and infant routinely use nonverbal forms of communication. For example, in one of the sessions, the program presents the "Still Face Experiment" by Tronick et al. (1978). The experiment involves both an infant child and his or her mother, who remains expressionless for three minutes. The finding from this experiment is that infants attempt many different forms of nonverbal communication (facial or gesture expressions) to elicit reactions from their mothers. The infants become upset with the mother's lack of response, and they cry or start self-soothing behaviors. The study remains one of the most replicated findings in developmental psychology, and it illustrates to economically diverse parents how infants use nonverbal communication in social settings.

<u>Coaching</u>: Each session leverages "Talking Tips" to allow parents to practice ways to improve the child's language environment without requiring significant changes in the parents' daily routines. Parents learn these talking tips by watching video vignettes that show other parents interacting with an infant or toddler. After observing other parents' behavior, the parents reflect on how they could use (with or without changes) these talking tips. As the program advances, the videos become more complex, and parents assess and discuss, as a group, which strategies other parents employed and whether parents in the vignettes missed any opportunities. When parents return to a later session, they are encouraged to share successful strategies, so they use their own experience to generate new talking tips that other parents in the group can also put into practice. Coaching models behavior and encourages parents to explore and identify new opportunities to improve the child's language environment.

The talking tips illustrate, in practice, vital aspects of a high-quality language environment: joint (or shared) attention, and speech recasting. The talking tips present easily accessible information for families about how children learn language from what they hear most, from things that interest them, from conversation and from positive relationships. Joint attention is the shared focus of two individuals on an object. An individual uses eye-gazing, pointing, or other verbal or non-verbal communication to initiate a session of joint attention with another individual. Events of joint attention promote language development because they provide a context that enables children to associate meaning to a particular utterance, thus promoting word comprehension and vocabulary expression (Bruner, 1983). For example, one of the tips is to "name things that the child is interested in." This tip reflects research that shows that infants are more likely to engage in joint attention when the parent refers to an object that the child is playing with, and far more difficult when the object is outside the child's attention (Rollins, 2003).

An event of speech recasting occurs when the adult repeats the child's speech with more detailed (or more correct) language. It allows the child to hear accurate language, thus helping the child figure out language syntax. The critical aspect of speech recasting is not to force the child to repeat after the adult, but rather for the adult to emphasize the linguistic feature the child needs to learn (Cleave et al., 1985). Indeed, one of the tips is for parents to "repeat and add to what they say and do." In summary, the LENA Start Program builds its "Talking Tips" component on the science that uncovers the contribution of joint attention and speech recasting to language development.

<u>Feedback</u>: At the end of each Group Session, each parent receives a LENA recorder, and each family completes a daylong recording with their children. When the parents return to the next LENA Start session, they hand in the recorder to the LENA Start Coordinator, who uploads the data and produces a feedback report. Parents are given a chance to review their report and notice the trends from one week to the next and the times of days of strength in communication. The LENA Start Coordinator then offers to discuss the report with each parent. The feedback follows a dialectical approach in which the Coordinators recognize that the parents are doing the best that they can, and, at the same time, points out areas in which the parents can improve. Therefore, when discussing the report, the Coordinator first praises the parents for their accomplishments (marked with stars). Second, the Coordinator identifies segments of the day in which the language environment still has some growth potential.

To be able to produce this feedback, the report summarizes conversational turns between adults and the focus child, adult words spoken around the focus child, which includes speech other than the speech directed to the child (e.g., a parent talking on the telephone), and the amount of time of exposure to electronics/TV, and self-reported reading minutes. The feedback report presents this information aggregated to the entire day (for the last eight recording days) and broken down in hourly fashion (for the last recording day).

Figure 1 displays a report template. The report contains three rows and two columns. The top row displays the percentiles (and counts) of adult words, the middle row shows the percentiles (and counts) of conversational turns, and the bottom row presents data about minutes of sounds from TV or other electronics. The left columns present these statistics aggregated by recording day for the last eight recording days. The right columns present the statistics broken down by the hour for the last recording day. The green text on the top right corner of the feedback report contains a parent self-report number of reading minutes. The feedback report contains the number of stars parents have received to date. A parent receives a star when they meet pre-specified targets for adult words, conversational turns, and reading minutes. The Coordinator praises the parent for the stars to recognize that they are doing the best they can.

Next, the Coordinator uses the report to identify opportunities for improvement. For example, when reviewing the report, the Coordinator will remark that the parent and the child engaged in much interaction (as measured by conversational turns by the hour, as reported in the middle row and right column of Figure 1) between 4 PM and 5 PM, but not so much at 11 AM (or 1 PM). The fact that the information is objective, detailed, and actionable may provide parents with opportunities to improve their child's language environment. It also changes the role of the coordinator to be a support and not the one passing judgement on a parent's skills. The primary focus is on providing feedback about conversational turns, and, if there is much exposure to TV/electronics, then the Coordinator will encourage parents to shift from TV to conversations with the child.

Book reading: At each session, the parents receive a children's books and are encouraged to read to their child. The book and ways to read and engage the child in conversation about the books are incorporated into each session. This part of the intervention aims to replicate the goals of programs such as "Reach Out and Read" (Weitzman et al., 2004).

Reporting: Approximately once a month, the parents answer the Developmental Snapshot (Gilkerson et al., 2017b). The Snapshot is an instrument designed to evaluate language skills for infants and toddlers. It provides developmental age and percentile ranking information compared to age-matched peers. The goal is for parents to start paying attention to their children's development and observing how their language skills have progressed. LENA Start Coordinators review the Developmental Snapshot data with the parent, and they screen for signals of severe language development delays that may require additional attention from an early intervention specialist.

3. Study Procedures

Next, we describe the study procedures approved by the University of Pennsylvania and Children's Hospital of Philadelphia Institutional Review Boards.

3.1 Eligibility for the LENA Start Program Evaluation Study Within the PHD Study

We used three criteria to determine eligibility for the LENA Start Program Evaluation Study. We restricted eligibility to PHD Study Participants in the inner-city sample. We restricted eligibility to mothers whose children were at most 33 months by May/2017 because the LENA Start Program is designed for children up to that age. Additionally, we attempted to recruit mothers who gave the research team authorization to be contacted for participation in future PHD Studies. We identified a group of 289 mothers who satisfied all of the three inclusion criteria.

3.2 Assignment to Control or Treatment Arms

After we identified the eligible families, and before we recruited them, we grouped PHD Study Participants according to the child's date of birth. We created ten blocks with 26 to 28 mothers in each group. Next, we took a random draw from a uniform distribution for each mother in each group. We created an "ordered invitation list" by ordering the mothers in descending order of the draw from the uniform distribution within each block. Let $z_i = 1$ if mother i is at the top half of the invitation list and $z_i = 0$ if the mother is at the bottom half of the list. In what follows, we refer to the mothers at the top half of the list (i.e., those with $z_i = 1$) as participants randomly assigned to the "treatment group" while mothers at the bottom half (those with $z_i = 0$) as study participants randomly assigned to the "control group."

3.3. Recruitment of Study Participants

The group format of the LENA Start Program creates specific challenges for the design of the evaluation protocol. The low-income families who participated in the PHD Study have unstable living arrangements or unforecastable work schedules. In such circumstances, recruitment strategies face a trade-off. On the one hand, the longer the recruitment period, the more likely we will locate an eligible family and consent the study's participants.[28] On the other hand, the longer the recruitment period, the more likely that parents who were recruited early on would report a change in the work schedule and could no longer attend group sessions at the time they had agreed to participate in recruitment (to the LENA Start Program). Our team decided to make an intense effort to recruit participants within three weeks and start the Program in the fourth week.[29]

Next, we describe the protocol for the recruitment of participants. We invited the vast majority of the mothers at the bottom half of the ordered invitation list to participate in the study's control arm.[30] Once invited, the mothers assigned to the control group could accept or decline our invitation. The parents who consented to participate in the evaluation study's control arm agreed to participate in the study's data collection procedures, which we explain below.

We invited the mothers to participate in the study's treatment arm at the top of the ordered invitation list. The mothers could accept or decline our invitation. If they accepted the invitation, the research assistants found a suitable schedule for the mothers to participate in the LENA Start sessions. If a mother assigned to the treatment group rejected our invitation to participate in the LENA Start Program, our research assistants then invited that mother to participate as a control group member. We did so for two reasons. First, we had a brief amount of time to recruit parents for the evaluation

[28] In the PHD Study, the second visit took place when the child was about 9 months old. We started efforts to locate the family when the child was 6 months old. The retention rate in the study was 80%.
[29] Group-based programs that serve low-income families can work better if they operate at a large scale in a central location that is easily accessible by families who depend on public transportation. The operation at a large scale would allow the program to be offered many times a day (from early in the morning to late evening), seven days a week, so that parents could attend sessions that best fit their schedule in any given week. Unfortunately, such a design is not feasible unless there is a stable and secure source of funding to run the program at large scale.
[30] To make sure we reached a minimum number of mothers in some groups, we invited six of the mothers at the bottom half of the ordered invitation list to participate in the LENA Start Program.

study. Because of our families' instability of living arrangements, we knew that it would not be feasible to contact a substantial fraction of our eligible study participants within three weeks. This recruitment protocol helped retain as many study participants as possible.

Second, we expected a low demand for the LENA Start Program. We formed our expectation from the literature that reports the demand for parenting education programs (e.g., Kalil, 2014). In studies that follow the typical protocol, they do not usually collect any data from parents who refuse to participate as the treatment group members. As a result, we usually know very little about such parents. Because they participate as members of the control group (although no longer in a randomized fashion), we can learn a little more about these parents with our team's protocol for this study. Also, we can still recover intent-to-treat treatment effect parameters because the assignment variable z_i is random.

3.4. Data Collection and Other Study Procedures

The parents who consented to participate in the study agreed to provide data on their children's language environment and answer a brief survey questionnaire. The procedures and instruments we used were identical for control and treatment groups and the same before (baseline) and after the program (endline).

3.4.1. *Collection of Language Environment Data*

We describe the logistics of the recording activities carried by families that participate in the study. The logistics we describe in this section relate to the data collection to evaluate the LENA Start Program. As such, we describe a protocol that applied to families in the study's control and treatment arms. Each parent received a LENA recorder and either a child's size vest or t-shirt that had a pocket for the recorder. Parents recorded the child's audio environment once at baseline and once at the endline. We asked the parent to insert the recorder in the clothing's front pocket in the morning, put the child's clothing, and let the recorder run for at least 12 hours. During this period, the device recorded the child's audio environment. We asked parents to record the language environment on a typical day for the child (e.g., not to record if the child was sick or an unusual event such as a birthday party).

After the family had finished the recording session, we retrieved the recording device and uploaded the audio file to a secure cloud server. The server contains software that processes the audio data and automatically produces four different statistics: number of conversational turns, number of child vocalizations, number of adult words, and minutes of audio from the TV or other electronics.[31] We focus our analysis on conversational turns and adult words in audio segments with at least one conversational turn between the focus child and an adult. We have standardized these variables to have mean zero and variance one in all of the estimates that we report below.

The LENA System identifies and labels individual adult and child utterances (also called segments), among other types of sounds. One conversational turn is defined as a speech-related vocal utterance initiated by the focus child or an adult to which the other responds within five seconds. Vocal utterances may include coos, squeals, babbles, and words. The number of conversational turns between the adults in the household and the child is the most critical measure of the language environment we use in our paper because it correlates with adult-child joint attention and speech recasting.[32] As summarized by Golinkoff et al. (2019), "back-and-forth conversations that are both temporally and topically contingent on the children's contribution are the fuel that prime the learning of language."

For adult utterances, the LENA System estimates the number of adult words spoken by any adult based on automated recognition of consonants and vowels and vocalization durations after filtering unclear speech. However, unlike adult word statistics commonly used in established research (e.g., Hart and Risley, 1995; Weisleder and Fernald, 2013), the LENA System estimate does not distinguish between words directed to the child and words overheard by the child. Research suggests that language learning from overheard speech does not occur reliably until children are about four or

[31] The estimates of the language environment variables LENA System produces are highly accurate and have been established in the literature (see Gilkerson and Richards, 2017).

[32] See Goldin-Meadow et al. (2014), Harris et al. (2010), Malin et al. (2014), McGillion et al (2017), Reed et al. (2017), Romeo et al. (2018), and Tamis-Lemonda et al. (2014) for additional insights on why conversational turns are so important for language development.

five years old (Rollins, 2003; Messenger, Yuan, and Fisher, 2015).[33] As explained by Golinkoff et al. (2018), overheard speech does not support early language environment because it demands a lot more attention from young children (they have to stop doing what they are doing and pay attention to what the other people are doing), it requires a level of social-cognitive skills that children may not have developed yet (because they need to understand the intentions of the adults involved in the conversation), and adult-directed speech is different from the child-directed speech in content, tone of voice, and grammatical complexity. For this reason, we use in our analysis adult words in audio segments that have at least one conversational turn with the focus child.

Finally, the LENA System identifies audio from the TV or other electronics as a separate category of the child's audio environment.

3.4.2. *Collection of Survey Data*

We interviewed the parents who agreed to participate in the study on two occasions. The baseline interview occurred just before the beginning of the 13-week LENA Start Program, and the endline interview took place immediately after the end of the program. Participation in the interview was high as all 136 parents who consented to participate in the study answered the survey questionnaire in the baseline, and 130 of the same parents did so in the endline.

We used the same survey questionnaire in the baseline and endline. The survey questionnaire had only eleven questions and focused on measuring four potential mechanisms that could help understand how program impacts (if any) were produced due to the program. First, we assessed parental beliefs about the importance of the language environment for a child's language development. The elicitation instrument has two items that elicit maternal beliefs about the importance of the language environment for a child's language development. These two items reflect scenarios of investments and ask mothers to predict the child's future language development. The first scenario describes a language environment with many conversational turns between adults and children, and little exposure to TV. In contrast, the second scenario represents a low-quality language

[33] It is possible that studies of overheard speech overestimate its impact on language development. The studies typically eliminate or at least greatly reduce other stimuli in the lab, so they minimize attentional demands from children. See, for example, Yuan and Fisher (2009).

environment, with lots of exposure to TV and little verbal exchange between adults and children. These hypothetical scenarios of language environment ("investments") follow the instrument design in Cunha, Elo, and Culhane (2013).

Second, we measured parental knowledge about the importance of the language environment for a child's language development. We used the instrument that Suskind et al. (2015) developed to evaluate a parenting program's impact on parental linguistic input and knowledge. The questionnaire has 30 items divided into five subscales. The first subscale covers parental perceptions about how children learn to talk.[34] The second subscale measures parental attitudes about reading to children.[35] The third subscale assesses parental perceptions about how children learn math.[36] The fourth subscale parental perceptions about the relationship between language development and school readiness.[37] The fifth and last subscale quantifies parental perceptions about TV and language development.[38] For each item, parents choose one alternative (among five) that describes the extent to which they agree with the statement. The alternative range from "Strongly Disagree" to "Strongly Agree."

Our survey questionnaire also measures parent self-efficacy (Bandura, 1977), which is defined as the parent's belief in their ability to perform the parenting role competently. The scale has four statements, and we ask participants to choose one (out of five) alternatives that range from "least sure" to "very sure." For example, one item states that "I know what my child should be able to do at each age as they grow." Coleman and Karraker (1998) and Jones and Prinz (2005) summarize the literature and report that parent self-efficacy correlates with child psychological functioning and child adjustment. They also found that parents with higher self-efficacy scores had higher levels of parenting competence and parenting satisfaction.

[34] The second item in this subscale states that: "Children learn fewer words when adults talk with a warm tone."
[35] The first item in this subscale states that "You cannot teach children anything new by reading them the same book over and over."
[36] The third item in this subscale states that "Talking about the difference between tall and short teaches toddlers about math.
[37] The third item in this subscale states that "How many words a two-year-olds know can predict how well they might do in kindergarten."
[38] The last item in this subscale states that "The more television children under two watch by themselves the more words they learn."

Finally, we ask parents to report their social support perception because of the intervention's group nature. The scale has four items, and we ask parents to choose one (out of five) alternatives that describe the extent to which they agree with the statement. For example, one statement is, “It is easy for me to talk with other parents about being a parent.” Previous research has shown that positive social support from family and friends increases parenting competence by providing encouragement and resources, particularly for first-time mothers (Leahy-Warren, McCarthy, & Corcoran, 2012).

3.4.3. The LENA Start Program in the PHD Study

The parents who consented to participate in the LENA Start Program agreed to attend the thirteen weekly sessions held at our Children’s Hospital of Philadelphia. Because of the group format, our team collected information from parents to find the times that suited the parents’ schedule. For the PHD Study participants, we learned that Saturday morning was particularly suitable for many parents. We also found that it was suitable to offer sessions one morning during a weekday. Because of space constraints, we attempted to have around ten parents at most. There are two attendance requirements that parents must fulfill to graduate from the LENA Start Program. First, the parent attended all of the first four sessions. Second, the parent had to complete at least five of the remaining nine sessions.[39]

4. Results

4.1 Recruitment for the LENA Start Program

In this section, we describe the results of our randomization procedure and recruitment efforts. Figure 2, as well as Tables 5 and 6, guide our discussion of the findings. We identified 289 PHD Study participants who satisfied all three inclusion criteria for participation in the LENA Start Program evaluation study. We compare the PHD Study participants that were eligible for participation in the LENA Start Evaluation Study with non-eligible ones.

The families residing in inner-city Philadelphia with a child who was 34 months or younger in May 2017 are eligible for the LENA Start Program. We randomly assigned the eligible families to

[39] If the parent could not attend a session, they could contact the LENA Start Coordinator to schedule a make-up session.

the control or invitation to the LENA Start Program arms. As shown in Table 5, the differences between the control and invitation groups are small and not statistically significant at the 5% level. However, we find that the counts of adult words are lower for the children in the invitation group, and that this difference has a p-value of 7.2%. In fact, we note that the LENA measures indicate a lower quality of language environment because the counts of conversational turns are higher, and exposure to TV is lower, for the control group children (even though the differences are not statistically significant).

Figure 2 describes the results of the randomization and recruitment procedures. We randomly assigned 145 of the 289 mothers in the Eligible Urban Group to the control group, and the remaining 144 to the group invited to participate in the LENA Start Program. In the three weeks of the recruitment effort, we managed to locate 94 (or 65%) of the 145 parents in the control group and 95 (or 66%) of the 144 parents in the LENA Start group's invitation.

Table 6 presents the results of our 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 probit model in which d_i^F is the dependent variable, and family characteristics are the explanatory variables. The explanatory variables include the assignment to control or invitation arms, a dummy variable for Hispanic ethnicity, a dummy variable for Non-Hispanic black, a dummy variable that takes the value one if the mother was born between 1978 and 1987. Mothers born in these years constitute the older group of mothers in the PHD Study, and the younger mothers were born in 1988 or later. We also include a dummy variable that is equal to one if the household income is less than or equal to two times the poverty line, the family's standardized HOME Score, and the child's Composite Language score at age two years. Column 1 in Table 6 shows that the assignment to control or invitation arms does not predict our team's success in locating a study participant. However, Table 6 also shows that our team was more likely to locate older, white, and lower-income mothers within the group of eligible study participants.

We return to Figure 2 to describe the results of the consenting procedures. Once we found the parents, we tried to consent the parents to participate in the study activities. Nearly 75% (71 out of 94) parents in the control group agreed to participate in the study. Approximately 68% (65 out of 95) of the treatment-group parents agreed to participate in the study. 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 probit model in which d_i^C is the dependent variable, and the family characteristics are the predicting variables. Column 2 in Table 6 shows that the most crucial variable in predicting whether a family consented to participate in the study or not was the random assignment to control or invitation arms. 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. Conditional on being located, low-income families had a greater likelihood of consenting to participate in the study.

We now return to Figure 2 to discuss attendance at the LENA Start Program. As we explained above, the acceptance to attend the LENA Start Program was not random because some of the families randomly assigned to receive the invitation declined the offer. Our team also invited a few members of the control group to have groups with ten families. As shown in Figure 2, the final acceptance list contained 61 families, 55 of which came from the study's invitation arm, and six of which came from the study's control arm. However, only 39 of the 61 parents attended at least one of the LENA Start Program sessions.

In our impact analysis below, we estimate two different treatment effect parameters. The first parameter is the "Intent to Treat" parameter (ITT) in which we use the binary variable z that captures random assignment to control ($z_i = 0$) or invitation arm ($z_i = 1$). Let $d_i = 1$ if the family attended the LENA Start Program and $d_i = 0$ otherwise. The second treatment effect parameter is the "Local Average Treatment Effect," which uses z_i as an instrumental variable for d_i. Column (3) in Table 6 shows the first stage in a two-stage least square procedure. We find that z_i is a strong instrument for d_i. Furthermore, because z_i is randomly assigned to individuals, we argue that z_i is a valid instrument for d_i.

4.2. Data Collection

4.2.1. *Survey Data*

As with the language environment data, we interviewed the parents who agreed to participate in the study on two occasions. Participation in the interview was high as all 136 parents who consented to participate in the study answered the survey questionnaire in the baseline, and 130 of the same parents did so in the endline.

4.2.2. *Language Environment 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. [40] A segment is defined to be valid if it satisfies four conditions. First, the segment was complete, meaning that the recording lasted precisely 300 seconds. About 1.6% of our recording segments were not complete, and we drop 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 and the second 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, and, for this reason, 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 lasted 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.

---

[40] 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, 104 parents for the endline. A total of 90 parents provided valid recording data for both baseline and endline recording sessions. Table 7 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. Columns (1) and (4) in Table 7 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.[41]

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. Although the OLS estimator's point estimates are not statistically significant, we find that the point estimate is large. Families randomly assigned to the invitation arm tend to provide 40% of a standard deviation longer for recording sessions. Once we account for selection, the coefficient turns statistically significant with virtually the same point estimate. Given the large discrepancy in recording length between the study's control and invitation arms, we control for recording length in all of our analyses below. In Appendix B, we also consider models with polynomials in recording length as control variables. We show that our results are not driven by differences in endline recording length between control and invitation arms.

Table 8 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:

$$Y_{i,0} = \beta_0 + \beta_1 Z_i + \sum_{j=1}^{J} \gamma_j G_{i,j} + \beta_3 RD_{i,0} + X_i \beta_5 + \delta_1 Sat_i + \delta_2 Sun_i + \epsilon_{i,0}. \quad (1)$$

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

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 variable $G_{i,j}$ takes the value one if family i was a member of the randomization group j, and zero otherwise. The variable $RD_{i,0}$ denotes the length of the recording session in the baseline data collection. The vector X_i captures demographic characteristics of family i. The dummy variables Sat_i and Sun_i take the value one if the recording took place on a Saturday and Sunday, respectively.

The error term $\epsilon_{i,0}$ is not necessarily independent across observations because of our study design. For this reason, we cluster observations by recruitment groups for all of our estimates of the standard errors. We believe that cluster is justified because programs will attempt to form groups of parents whose children are of similar ages to encourage parents to share experience and exchange ideas about implementing the "Talking Tips."

The findings are the same across the four models. There are small differences in the number of conversational turns between the study's control and invitation arms at baseline. Although the point estimates suggest are not statistically significant, 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 if there are pre-existing differences between parents who accept the invitation and participate in the LENA Start Program or not. The set of parents who attend at least one of the LENA Start Program sessions is a potentially selected set of all of the parents who were invited. 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:

$$D_i = \alpha_0 + \alpha_1 Z_i + \sum_{j=1}^{J} \pi_j G_{i,j} + \alpha_3 RD_{i,0} + X_i \alpha_5 + \theta_1 Sat_i + \theta_2 Sun_i + \omega_{i,0} \tag{2}$$

$$Y_{i,0} = \beta_0 + \beta_1 D_i + \sum_{j=1}^{J} \gamma_j G_{i,j} + \beta_3 RD_{i,0} + X_i \beta_5 + \delta_1 Sat_i + \delta_2 Sun_i + \epsilon_{i,0} \tag{3}$$

We argue that the random assignment to control or invite the study's arms is a valid instrument because it satisfies both the exclusion restriction and a strong first stage (see Table 6). We find that the 2SLS estimator of β_1 is not statistically significant for three models, but it is statistically significant (at the ten percent level) for the Model 4.

Most of the estimates of β_1 in Table 8 are not statistically significant. 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 take into account these pre-existing differences.

4.3. Effects of the LENA Start Program on Conversational Turn Counts

Next, we turn our attention to the analysis of the endline recording data. We start our presentation with a focus on conversational turns because it is our primary variable of interest. Panel A in Table 5 presents the OLS estimators of β_1 and β_2 of variations of the following regression model:

$$Y_{i,1} = \beta_0 + \beta_1 Z_i + \sum_{j=1}^{J} \gamma_j G_{i,j} + \beta_2 Y_{i,0} + \beta_3 RD_{i,1} + \beta_4 RD_{i,0} + X_i \beta_5 + \delta_1 Sat_{i,1} + \delta_2 Sun_{i,1} + \delta_3 Sat_{i,0} + \delta_4 Sun_{i,0} + \epsilon_{i,1}. \quad (4)$$

We consider three models. The first model imposes $\beta_2 = \beta_4 = \beta_5 = \delta_3 = \delta_4 = 0$. Thus, it does not include the lagged outcome or the recording duration at baseline, and it does not include covariates for demographic characteristics. The subsequent models include each one of these components, one at a time. In the second model, we include demographic characteristics (so we relax the restriction that $\beta_5 = 0$). In the third model, we add the conversational turn counts at baseline, the recording duration at baseline, and dummies for Saturday and Sunday at baseline. Therefore, we estimate all of the coefficients in (4) in the fourth specification.

Table 9 reports mixed findings for the impact of the LENA Start Program on conversational turn counts. According to the OLS estimator of Model 1, the LENA Start Program increases conversational turns between adults and the focus child by over 13% of a standard deviation, but it is not statistically significant. When we control for demographic characteristics, the point estimate of the program's intent-to-treat effect (ITT) increases to 21.5% of a standard deviation, but it still is statistically insignificant. In model 3, when we control for conversational turns at baseline (and variables that describe the recording characteristics at baseline), the ITT's point estimate is 31.4% of a standard deviation, and it is statistically significant at a 5% confidence level.

We now turn to the analysis of the impacts of the LENA Start Program on parents who attended at least one session. To account for selectivity in attendance, Panel B in Table 5 presents the results from the following two-stage least squares regression model:

$$D_i = \alpha_0 + \alpha_1 Z_i + \sum_{j=1}^{J} \pi_j G_{i,j} + \alpha_2 Y_{i,0} + \alpha_3 RD_{i,1} + \alpha_4 RD_{i,0} + X_i \alpha_5 + \theta_1 Sat_{i,1} + \theta_2 Sun_{i,2} + \theta_3 Sat_{i,0} + \theta_4 Sun_{i,0} + \omega_{i,1} \quad (5)$$

$$Y_{i,1} = \beta_0 + \beta_1 D_i + \sum_{j=1}^{J} \gamma_j G_{i,j} + \beta_2 Y_{i,0} + \beta_3 RD_{i,1} + \beta_4 RD_{i,0} + X_i \beta_5 + \delta_1 Sat_{i,1} + \delta_2 Sun_{i,1} + \delta_3 Sat_{i,0} + \delta_4 Sun_{i,0} + \epsilon_{i,1} \quad (6)$$

According to Model 1, the program's impact on conversational turns, as estimated by the LATE parameter, is 23% of a standard deviation. If we control for demographic characteristics, the LATE increases to 38.3% of a standard deviation, and it is statistically significant at the 10% level. If we control for conversational turn counts at baseline, the LATE increases to 55.1% of a standard deviation, and it is statistically significant at the 1% level.

To put the findings in a broader context, we simulate equation (6) in the following way. For all individuals, we eliminate variability in conversational turns due to recording duration and fix $RD_{i,1}$ (before standardization) to twelve hours. For all of the other variables, we use mean values for the estimation sample. 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:

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

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

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 (2017), 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 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 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:

$$Y_{i,t} = \beta_0 + \beta_1 Z_{i,t} + \beta_2 t + \beta_3 RD_{i,t} + \beta_4 X_{i,t} + \delta_1 Sat_{i,t} + \delta_2 Sun_{i,t} + \eta_i + \epsilon_{i,t}. \quad (7)$$

In specification (7), 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 (7). As we show in Table 10, 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 1% level.

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:

$$D_{i,t} = \alpha_0 + \alpha_1 Z_{i,t} + \alpha_2 t + \alpha_3 RD_{i,t} + \alpha_4 X_{i,t} + \theta_1 Sat_{i,t} + \theta_2 Sun_{i,t} + \eta_i + \omega_{i,1} \quad (8)$$

$$Y_{i,t} = \beta_0 + \beta_1 D_{i,t} + \beta_2 t + \beta_3 RD_{i,t} + \beta_4 X_{i,t} + \delta_1 Sat_{i,t} + \delta_2 Sun_{i,t} + \eta_i + \epsilon_{i,t}. \quad (9)$$

We present the results of the IV-FE model in Table 10. As estimated by the IV-FE procedure, the program's impact is 57.8% of a standard deviation, and it is statistically significant at the 1% 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.

These large impacts may arise because parents in the treatment group converse a lot more with the children but only when they are wearing the LENA recorder. It would be challenging to design and implement an ethical study that would comprehensively address such legitimate criticism because any ethical study would have to disclose all of the study procedures (including recording the child's language environment). If we assume that the children – because they are toddlers – do not change their language behavior because of the recorder's presence, we have an indirect way of assessing whether the improvements in conversational turns are just for show or reflect real changes in interactions between parents and children. This indirect way consists of using the LENA Advanced-Data Extractor (ADEX) to break the audio file into segments that include conversations between the focus child and an adult (male or female). We can then identify who initiated the conversation (the child or a male or a female adult), and the number of conversation turns per segment as categorized by the segment initiator (the child, a female adult, or a male adult). If the adults change their behavior because of the recorder (and if the toddlers are not), then the differences in conversation turn between the control and treatment groups will reflect the differences in conversations initiated by the (female or male) adults.

On the other hand, if the adults apply what they have learned from the LENA Start Program, they will let the child take the lead (which is one of the fourteen "Talking Tips"). In this case, the changes in conversational turns arise because the children in the treatment group initiate more conversations than their control counterparts, and the adults follow the child's lead and respond in ways that require a response from the child. Therefore, both the number of conversations (the number of continuous audio segments with the focus child and an adult) and the number of conversational turns increase because of the change in the parental responses to an initiation of a focus child's conversation.

We use the ADEX audio segmentation tool to investigate whether children or adults initiate more conversations (audio segments). For this analysis, we use the lagged variable model (4) to estimate the ITT parameter, and the model described in equations (5) and (6) to estimate the LATE parameter. Panel A in Table 11 presents the results of our analysis of the initiation of segments of conversation. Both the OLS and the 2SLS estimators show that the differences in conversational turns arise because the LENA Start Program children initiate a more significant number of conversations. According to the ITT, the treatment group children initiate nearly 40% of a standard deviation more conversational turns than their control counterparts. The LATE parameters suggest that the impact is 69% of a standard deviation. The smallest impact group are female adults, and all of the LENA Start participants (except for one) were female adults.

Panel B reports the number of conversational turns in segments initiated by the focus child, a female adult, and a male adult. The ITT and the LATE parameters show that the changes in conversations initiated by the focus child drive conversational turns. We find suggestive evidence (for the LATE) that male adults increase the number of conversational turns with children, even though there was only one male LENA Start Program participant. If ever replicated, it suggests spillover within families.

These findings show that the LENA Start parents do not initiate more conversations than the control group's parents. However, LENA Start children do initiate more conversations, and, when they do so, the LENA Start parents are responding in ways that lead to a more significant number of conversational turns. These findings reduce (but do not eliminate) the concern that the increase in conversational turns occurs when they are wearing the device.

Our findings are significant because of the existing research on the differences between high- and low-quality language environments. Hart and Risley (1995) document that the disparity in words addressed to children results from children in higher-income households initiating more conversations and adults responding in ways that lead to more conversational turns and, as a result, more words directed to the child. This finding suggests that the language environment is more affluent because parents are talking about objects of interest to the child, and thus the conversations occur during joint attention events.

Before we turn to the analysis of the impact of the program on adult words, we briefly remark on our investigation of the robustness of our point estimates. First, we investigate the sensitivity of the ITT estimates with respect to 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 proposed in Lee (2009), which also allow for endogenous attrition. Our results are not sensitive to attrition. The IPW procedure produces a point estimate for the ITT that is slightly lower (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 not possible to reject the null (no impact) for the lower bound, but the confidence interval for the upper bound does not include zero. We report these findings in Appendix Table B1.

Second, we also investigate how the ITT and LATE changes once we control for recording duration in flexible ways. As we show in Appendix Table B2, the point estimates are robust to the way we control for recording duration. If anything, the standard errors or our estimates of the program impacts decrease as we specify recording duration in more flexible ways.

In what follows, we investigate the program's impact on adult words in audio segments with a conversation between the focus child and an adult.

4.4. Effects of the LENA Start Program on Adult Words

Next, we consider 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 both words addressed to the child as well as 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 (4) and (5)-(6), but the results are the same if we use the FE or IV-FE procedures described in equations (7) and (8)-(9).

Table 12 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 both 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 in turns. 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 attendant was a male, and all the others were female. This finding reinforces the evidence of spillover of the program to the male adult in the household.

Panel B in Table 12 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 by parents in ways that do not stop the conversation, but rather allow the child to continue to talk and interact.

4.5. Effects of the LENA Start Program on Exposure to TV/Electronics

We also investigated the impact of the LENA Start Program on the focus child exposure to audio from TV or other electronics. Table 13 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, then the coefficient becomes even smaller but it is still is statistically insignificant. We conclude that the LENA Start Program does not produce a decrease in exposure to TV/Electronics.

4.6. Mechanisms

Finally, we study the potential mechanisms of the impacts of the program. As described in Section 4.4.2., we measured four potential mechanisms: parental beliefs, parental knowledge, parental self-efficacy, and parents' report of social support. Appendix A contains detailed information about how we operationalize these constructs.

Table 14 presents the results of our analysis. As in our analysis of the impacts of the LENA Start Program on the language environment, we use the lagged dependent variable model described in equations (4) and (5)-(6) in our analysis. First, Columns 1 and 2 show the impacts on the other two possible mechanisms for the program's impact. The ITT or LATE estimators find null impacts of the LENA Start Intervention on parents' self-efficacy or parents' sense of social support. The point estimates are mostly small and close to zero. These findings rule out self-efficacy and social support as likely mechanisms for the program's impacts.

Column 3 presents the ITT and LATE estimates of the impact of the LENA Start Program on parental beliefs. The ITT's point estimate is about 32% of a standard deviation, while the LATE is about 61% of a standard deviation. The former is statistically significant at the 10% level, and the latter at the 5% level. The effect sizes on parental beliefs are comparable to the effect sizes on conversational turns.

Column 4 shows the ITT and LATE estimates of the impact of the LENA Start Program on parental knowledge. The impacts are similar in effect sizes, but the ITT is not statistically significant.[42] In Appendix A, we show that the aggregate result for the parental knowledge scale masks heterogeneous impacts by subscales.

Finally, we estimate, via two-stage least squares, the relationship between parental beliefs and the conversational turn counts. We use the random assignment as an instrument for parental beliefs. A problem of conducting inference is because Z_i may be a weak instrument. For this reason, we report both the Wald and Anderson-Rubin confidence intervals. The former is not robust to weak instruments, but the second is. We find that the first-stage coefficient, that is, the coefficient on the instrumental variable Z_i, has a point estimate around .406, a t-statistic of 2.40, and p-value equal to 0.019. Therefore, weak IV is a possibility. The second-stage estimate (i.e., the coefficient on parental beliefs) has point estimates around 0.74, a t-statistic of 1.97, and p-value equal to 0.048. The Wald confidence interval is [.005267, 1.47392], while the Anderson-Rubin interval is [.250043, 2.20825]. Therefore, both confidence intervals are bounded away from zero. This evidence reinforces our interpretation that the LENA Start Program improves parental linguistic input by shifting parental beliefs about the importance of early investments for early childhood development.

Conclusion

The LENA Start Program increases parent-child conversational turns and adult words per turn. These findings matter because parental linguistic input correlates with early language processing skills (Weisleder & Fernald, 2013), vocabulary growth (Pan et al., 2005), higher language development in early adolescence (Gilkerson et al., 2018), and more robust activation of the Broca's area of the brain (Romeo et al., 2018), which, in turn, are linked to the development of reading skills (Foorman et al., 2018). Differences in parental response to conversations initiated by the child drive the impacts of the program.

We tested four different mechanisms. We found no evidence in favor of parental self-efficacy or parent social support. We found suggestive evidence that the program's impact is due to improvements in parental beliefs and parental knowledge about the importance of early language environment for a child's language development.

[42] We also investigated the robustness of our findings by estimating fixed-effect models and two-stage least squares fixed-effect models. The point estimates are smaller and the standard errors tend to be larger. Therefore, the treatment effect parameters are not statistically significantly different from zero, but the point estimates on the impacts of the LENA Start Program on maternal beliefs are, again, large and imprecisely estimated.

References

Attanasio, O., Cunha, F., & Jervis, P. (2019). *Subjective Parental Beliefs: Their Measurement and Role*. NBER Working Paper 26516.

Baker, L., Mackler, K., Sonnenschein, S., & Serpell, R. (2001). Parents' Interactions with Their First-Grade Children During Storybook Reading and Relations with Subsequent Home Reading Activity and Reading Achievement. *Journal of School Psychology, 39*(5), 415-438.

Bandura, A. (1977). Self-Efficacy - toward a Unifying Theory of Behavioral Change. *Psychological Review, 84*(2), 191-215. doi:Doi 10.1037//0033-295x.84.2.191

Boneva, T., & Rauh, C. (2018). Parental beliefs about returns to educational investments—the later the better? *Journal of the European Economic Association,, 16*(6), 1669-1711.

Bown, L. (1990). *Repairing the Future: Women, Literacy, and Development. The Impact of Female Literacy on Human Development and the Participation of Literate Women in Change* Retrieved from Somerset, England:

Bradley, R. H., & Caldwell, B. M. (1980). The Relation of Home-Environment, Cognitive Competence, and Iq among Males and Females. *Child Development, 51*(4), 1140-1148. doi:DOI 10.1111/j.1467-8624.1980.tb02664.x

Bruner, J. (1983). *Child's talk: Learning to use language.* . New York, NY: W.W. Norton & Co.

Burchfield, S., Hua, H., Baral, D., & Rocha, V. (2002). *A Longitudinal Study of the Effect of Integrated Literacy and Basic Education Programs on Women's Participation in Social and Economic Development in Nepal.* Washington, DC: United States Agency for International Development.

Canieso-Doronila, M. L. (1996). *Landscapes of Literacy: An Ethnographic Study of Functional Literacy in Marginal Philippine Communities*. Hamburg, Germany: UNESCO Institute for Education and Luzac Oriental.

Carr-Hill, R. A., Okech, A., Katahoire, A. R., Kakooza, T., Ndidde, A. N., & Oxenham, J. (2001). *Adult Literacy Programs in Uganda*. Washington, DC: Human Development, Africa Region, World Bank.

Chebanne, A., & Nyati-Ramahobo, L. (2003). The development of minority language for adult literacy in Botswana: towards cultural diversity. *Southern Africa Review of Education with Education With Production, Volume 9*(1), 75-93.

Christakis, D. A., & Zimmerman, F. J. (2006). *The Elephant In The Living Room: Make Television Work for Your Kids*. New York, NY: Rodale Books.

Cleave, P. L., Becker, S. D., Curran, M. K., Van Horn, A. J. O., & Fey, M. E. (2015). The efficacy of recasts in language intervention: a systematic review and meta-analysis *American journal of speech-language pathology, 24*, 237-255.

Coleman, P. K., & Karraker, K. H. (1998). Self-efficacy and parenting quality: Findings and future applications. *Developmental Review, 18*(1), 47-85. doi:10.1006/drev.1997.0448

Cunha, F., Elo, I., & Culhane, J. (2013). *Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation*. NBER Working Paper 19144, (19144).

Cunha, F., Elo, I., & Culhane, J. (2020). Maternal subjective expectations about the technology of skill formation predict investments in children one year later. *Journal of Econometrics, In Press*.

Cunha, F., Heckman, J. J., Masterov, D., & Lochner, L. (2006). Interpreting the Evidence on Life Cycle Skill Formation. In E. Hanushek & F. Welch (Eds.), *The Handbook of Economics of Education* (Vol. 1, pp. 697-812). Amsterdam: North Holland.

Dexter, E. R., LeVine, S. E., & Velasco, P. M. (1998). Maternal schooling and health-related language and literacy skills in rural Mexico. *Comparative Education Review, 42*(2), 139-162. doi:10.1086/447493

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech Language and Hearing Research, 42*(6), 1432-1443. doi:10.1044/jslhr.4206.1432

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

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test | Fourth Edition*. San Antonio, TX.

Education, F. D. o. (2009–2014). *Florida Assessments for Instruction in Reading K-2 Technical Manual*. Tallahassee, FL Retrieved from http://www.fcrr.org/FAIR/Final_K-2_Technical%20Manual_2010.pdf

Egbo, B. (2000). *Gender, Literacy and Life Chances in Sub-Saharan Africa* Cleveland/Buffalo/Sydney: Multilingual Matters.

Fernald, A., Marchman, V. A., & Weisleder, A. (2012). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science, 16*(2), 234-248.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In *Developmental psycholinguistics: On-line methods in children's language processing.* (pp. 97-135). Amsterdam, Netherlands: John Benjamins Publishing Company.

Foorman, B. R., Petscher, Y., & Herrera, S. (2018). Unique and common effects of decoding and language factors in predicting reading comprehension in grades 1–10. *Learning and Individual Differences, 63*(1), 12-23.

Gilkerson, J., & Richards, J. A. (2008). *The LENA Natural Language Study*. Retrieved from https://www.lena.org/wp-content/uploads/2016/07/LTR-02-2_Natural_Language_Study.pdf

Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018). Language Experience in the Second Year of Life and Language Outcomes in Late Childhood. *Pediatrics, 142*(4). doi:ARTN e20174276 10.1542/peds.2017-4276

Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W., & Small, S. L. (2014). New Evidence About Language and Cognitive Development Based on a Longitudinal Study Hypotheses for Intervention. *American Psychologist, 69*(6), 588-599. doi:10.1037/a0036886

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. Language Matters: Denying the Existence of the 30-Million-Word Gap Has Serious Consequences. *Child Development, 0*(0). doi:10.1111/cdev.13128

Gough, P. B., & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education, 7*(1), 6-10. doi:10.1177/074193258600700104

Hannum, E., & Buchmann, C. (2003). *The consequences of global educational expansion: Social science perspectives*. Cambridge, MA: American Academy of Art and Sciences.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review, 73*, 103-130. doi:10.1016/j.euroecorev.2014.10.006

Harris, J., Golinkoff, R. M., & Hirsh-Pasek, K. (2010). Lessons from the crib for the classroom: How children really learn vocabulary. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of Early Literacy Research* (Vol. 3). New York, NY: Guilford.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes Publishing Company.

Hirotani, M., Stets, M., Striano, T., & Friederici, A. D. (2009). Joint attention helps infants learn new words: event-related potential evidence. *Neuroreport, 20*(6), 600-605. doi:10.1097/WNR.0b013e32832a0a7c

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*(5), 1368-1378. doi:10.1111/1467-8624.00612

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*(1), 55-88. doi:10.1016/j.dr.2005.11.002

Huttenlocher, J., Vasilyeva, M., Waterfall, H. R., Vevea, J. L., & Hedges, L. V. (2007). The varieties of speech to young children. *Developmental Psychology, 43*(5), 1062-1083. doi:10.1037/0012-1649.43.5.1062

Huttenlocher, J., Waterfall, H. R., Vasilyeva, M., Vevea, J. L., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology, 61*(4), 343-365. doi:10.1016/j.cogpsych.2010.08.002

Ireland, T. D. (1994). Literacy skills as building bricks for trade union democracy: the experience of the construction workers trade union in Joao Pessoa, Brazil. *Adult Education and Development, 42*(81-87).

Jones, T. L., & Prinz, R. J. (2005). Potential roles of parental self-efficacy in parent and child adjustment: A review. *Clinical Psychology Review, 25*(3), 341-363. doi:10.1016/j.cpr.2004.12.004

Kagitcibasi, C., Goksen, F., & Gulgoz, S. (2005). Functional adult literacy and empowerment of women: Impact of a functional literacy program in Turkey. *Journal of Adolescent & Adult Literacy, 48*(6), 472-489. doi:10.1598/Jaal.48.6.3

Kalil, A. (2014). Inequality Begins at Home: The Role of Parenting in the Diverging Destinies of Rich and Poor Children. In P. R. Amato, A. Booth, S. M. McHale, & J. Van Hook (Eds.), *Families in an Era of Increasing Inequality: Diverging Destinies* (pp. 63-82). New York: Springer.

Leahy-Warren, P., McCarthy, G., & Corcoran, P. (2012). First-time mothers: social support, maternal parental self-efficacy, and postnatal depression. *Journal of Clinical Nursing, 21*(3-4), 388-397. doi:10.1111/j.1365-2702.2011.03701.x

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies, 76*(3), 1071-1102.

Leech, K. A., Wei, R., Harring, J. R., & Rowe, M. L. (2018). A brief parent-focused intervention to improve preschoolers' conversational skills and school readiness *Developmental Psychology, 54*(1), 15-28.

List, J. A., Pernaudet, J., & Suskind, D. L. (2020). *It All Starts with Beliefs: Addressing the Roots of Educational Inequities by Changing Parental Beliefs*.

Malin, J. L., Cabrera, N. J., & Rowe, M. L. (2014). Low-income minority mothers' and fathers' reading and children's interest: Longitudinal contributions to children's receptive vocabulary skills. *Early Childhood Research Quarterly, 29*(4), 425-432. doi:10.1016/j.ecresq.2014.04.010

McDowell, K. D., Lonigan, C. J., & Goldstein, H. (2007). Relations among socioeconomic status, age, and predictors of phonological awareness. *Journal of Speech Language and Hearing Research, 50*(4), 1079-1092. doi:10.1044/1092-4388(2007/075)

McGillion, M., Herbert, J. S., Pine, J., Vihman, M., dePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What Paves the Way to Conventional Language? The Predictive Value of Babble, Pointing, and Socioeconomic Status. *Child Development, 88*(1), 156-166. doi:10.1111/cdev.12671

McGillion, M., Pine, J. M., Herbert, J. S., & Matthews, D. (2017). 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*(10), 1122-1131.

Messenger, K., Yuan, S., & Fisher, C. (2015). Learning Verb Syntax via Listening: New Evidence From 22-Month-Olds. *Language Learning and Development, 11*(4), 356-368. doi:10.1080/15475441.2014.978331

Morisset, C. E., Barnard, K. E., Greenberg, M. T., Booth, C. L., & Spieker, S. J. (1990). Environmental influences on early language development: The context of social risk. *Development and Psychopathology, 2*(2), 127-149.

Murnane, R. J., Willett, J. B., Duhaldeborde, Y., & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis* *and Management, 19*(4), 547-568. doi:Doi 10.1002/1520-6688(200023)19:4<547::Aid-Pam2>3.0.Co;2-#

Norwood, G. (2003). *Fighting oppression through literacy: a case study of the Karen's Women's Organisation*. ASPBAE Indigenous Adult Education Case Study Series. Mumbai, India.

OECD. (2016). *Skills Matter: Further Results from the Survey of Adult Skills*. Retrieved from Paris:

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76*(4), 763-782. doi:DOI 10.1111/j.1467-8624.2005.00876.x

Paslawski, T. (2005). The Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4):A Review. *Canadian Journal of School Psychology, 20*(1-2), 129-134. doi:10.1177/0829573506295465

Purcell-Gates, V., & Waterman, R. A. (2000). *Now We Read, We See, We Speak: Portrait of Literacy Development in an Adult Freirean-Based Class*. Mahwah, NJ: Lawrence Erlbaum.

Ramirez, N. F., Lytle, S. R., & Kuhl, P. K. (2020). Parent coaching increases conversational turns and advances infant language development. *Proceedings of the National Academy of Sciences of the United States of America, 117*(7), 3484-3491.

Reed, J., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Learning on Hold: Cell Phones Sidetrack Parent-Child Interactions. *Developmental Psychology, 53*(8), 1428-1436. doi:10.1037/dev0000292

Robinson-Pant, A. (2000). Women and literacy: a Nepal perspective. *International Journal of Educational Development, 20*(4), 349-364. doi:Doi 10.1016/S0738-0593(99)00080-2

Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics, 24*(2), 221-234. doi:10.1017/S0142716403000110

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. E. (2018). Beyond the 30-Million-Word Gap: Children's Conversational Exposure Is Associated With Language-Related Brain Function. *Psychological Science, 29*(5), 700-710. doi:10.1177/0956797617742725

Rowe, M. L. (2012). A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development. *Child Development, 83*(5), 1762-1774. doi:10.1111/j.1467-8624.2012.01805.x

Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science, 12*(1), 182-187. doi:10.1111/j.1467-7687.2008.00764.x

Rowe, M. L., & Leech, K. A. (2019). A parent intervention with a growth mindset approach improves children's early gesture and vocabulary development. *Developmental Science, 22*, e12792.

Sabatini, J., Bruce, K., & Steinberg, J. (2013). *SARA Reading Components Tests, RISE Form: Test Design and Technical Adequacy*. Retrieved from

Sandiford, P., Cassel, J., Montenegro, M. M., & Sanchez, G. (1995). The Impact of Women's Literacy on Child Health and Its Interaction with Access to Health-Services. *Population Studies-a Journal of Demography, 49*(1), 5-17. doi:10.1080/0032472031000148216

Schieffelin, B. B., & Ochs, E. (1986). Language Socialization. *Annual Review of Anthropology, 15*, 163-191. doi:DOI 10.1146/annurev.an.15.100186.001115

Suskind, D. L., Leffel, K. R., Hernandez, M. W., Sapolich, S. G., Suskind, E., Kirkham, E., & Meehan, P. (2013). An Exploratory Study of "Quantitative Linguistic Feedback": Effect of LENA Feedback on Adult Language Production. *Communication Disorders Quarterly, 34*(4), 199-209.

Talks, P. (2015). *Providence Talks: Pilot findings and next steps*. Retrieved from http://www.providencetalks.org/wp-content/uploads/2015/10/Providence-Talks-Pilot-Findings-Next-Steps.pdf

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. L. (2014). Why Is Infant Language Learning Facilitated by Parental Responsiveness? *Current Directions in Psychological Science, 23*(2), 121-126. doi:10.1177/0963721414522813

Tarawa, N. (2003). *Reading our Land: a Case Study of Te Waka Pu Whenua, Maori Adult Education Center in New Zealand* ASPBAE Indigenous Adult Education Case Study Series. Mumbai, India.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2011). *Test of word reading efficiency (2nd ed)*. Austin, TX.

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The Infant's Response to Entrapment between Contradictory Messages in Face-to-Face Interaction. *Child and Adolescent Psychiatry, 17*(1), 1-13.

UNESCO. (1975). International-Symposium for Literacy Declaration of Persepolis. *Convergence, 8*(4), 5-10. Retrieved from <Go to ISI>://WOS:A1975BK17900002

Vernon-Feagans, L., Hammer, C. S., Miccio, A., & Manlove, E. E. (2001). Early language and literacy skills in low-income African American and Hispanic children. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (Vol. 1). New York, NY: Guilford Publications.

Vignoles, A., De Coulon, A., & Marcenaro-Gutierrez, O. (2011). The value of basic skills in the British labour market. *Oxford Economic Papers-New Series, 63*(1), 27-48. doi:10.1093/oep/gpq012

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2012). *Comprehensive test of phonological processing (2nd ed.)*. Austin, TX.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children, 5th Ed.* . Bloomington, MN: Pearson.

Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science, 24*(11), 2143-2152. doi:10.1177/0956797613488145

Weitzman, C. C., Roy, L., Walls, T., & Tomlin, R. (2004). More Evidence for Reach Out and Read: A Home-Based Study. *Pediatrics, 113*(5), 1248-1253.

Williams, K. T. (2007). *EVT-2: Expressive Vocabulary Test, 2nd Ed.* Bloomington, MN: Psychcorp.

Yuan, S., & Fisher, C. (2009). "Really? She Blicked the Baby?": Two-Year-Olds Learn Combinatorial Facts About Verbs by Listening. *Psychological Science, 20*(5), 619-626. doi:10.1111/j.1467-9280.2009.02341.x

Zhang, Y., Xu, X., Jiang, F., Gilkerson, J., Xu, D., Richards, J. A., . . . Topping, K. (2015). Effects of quantitative linguistic feedback to caregivers of young children: A pilot study in China. *Communication Disorders Quarterly, 37*(1), 16-24. doi:10.1177/1525740115575771

Tables and Figures

Table 1
Characteristics of PHD Study by Residence

	Inner-City Sample (1)		Suburban Sample (2)		p-value (2) - (1)
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)	0.000
Mother is married	657	0.180 (0.380)	165	0.790 (0.410)	0.000
Standardized HOME Score	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 of Infant Development	541	-0.561 (0.740)	133	0.213 (0.818)	0.000

Note: Standard errors in parenthesis.

Table 2
Between and Within Sum of Squares as Fractions of Total Sum of Squares
Inner-City and Suburban Samples
PHD Study Data

Variable	Between	Within
Standardized HOME Score	11.2%	88.8%
Conversational Turn Counts (12 hours)	1.3%	98.7%
Standardized BSID Language Composite Score	14.3%	85.7%

Table 3
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***	0.093
	(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)
Constant	-0.442***	-1.938***
	(0.140)	(0.478)
Observations	234	239
R-squared	0.196	0.094

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 4
Correlation between Language Development with HOME and LENA Measures
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)
	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)
	Panel C: Standardized Adult Word Counts Only			
Standardized Adult Word Counts	0.100*	0.102*	0.101*	0.088
y	(0.059)	(0.054)	(0.054)	(0.055)
	Panel D: Standardized TV Time			
Standardized TV Time	-0.170***	-0.013	-0.002	-0.004
	(0.059)	(0.057)	(0.057)	(0.057)
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)
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

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

Table 5
Characteristics of PHD Study by LENA Start Assignment

	Invitation to LENA Start		Control		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	142	0.704 (0.458)	145	0.752 (0.434)	0.366
Mother has attended some higher education program	144	0.097 (0.297)	145	0.103 (0.306)	0.860
Mother is Hispanic	144	0.153 (0.361)	145	0.200 (0.401)	0.293
Mother is Non-Hispanic black	144	0.625 (0.486)	145	0.579 (0.495)	0.428
Mother is Non-Hispanic white	144	0.139 (0.347)	145	0.159 (0.367)	0.637
Mother is single	144	0.729 (0.446)	145	0.655 (0.477)	0.173
Mother is cohabiting	144	0.063 (0.243)	145	0.083 (0.276)	0.507
Mother is married	144	0.208 (0.408)	145	0.262 (0.441)	0.282
Standardized HOME Score	133	-0.204 (1.383)	132	-0.044 (0.887)	0.338
Conversational Turns at 9 months	47	283.279 (191.05)	54	317.517 (120.90)	0.265
Adult Word Counts at 9 months	47	13761.77 (6981.44)	54	16260.48 (6804.08)	0.072
Seconds of Exposure to TV	47	7308.39 (5007.89)	54	7052.62 (4813.34)	0.795
Standardized Language Score from the Bayley Scale of Infant Development	137	-0.443 (0.71)	134	-0.502 (0.82)	0.521

Note: Standard error in parenthesis.

Table 6
Family Location, Family Consenting, and Family Attendance to LENA Start Program

VARIABLES	(1) Family was located	(2) Family consented conditional on being located	(3) Family attended LENA Start Program conditional on consenting
Random assignment to LENA Start Program	-0.061 (0.160)	-0.790*** (0.274)	2.137*** (0.608)
Mother was born between 1978 and 1987	0.846*** (0.189)	-0.004 (0.265)	0.652* (0.345)
Mother is Hispanic	-1.186*** (0.326)	-1.346** (0.557)	-0.053 (0.534)
Mother is Non-Hispanic black	-1.117*** (0.367)	-0.885 (0.586)	0.757 (0.480)
Mother is cohabiting	-0.373 (0.407)	-0.280 (0.788)	-0.298 (0.626)
Mother is married	-0.899*** (0.279)	-0.402 (0.400)	0.102 (0.661)
Mother has some post-secondary education	-0.465** (0.189)	-0.642* (0.384)	0.147 (0.570)
Family is poor	0.900*** (0.330)	0.984*** (0.248)	-0.299 (0.474)
Standardized HOME Score	0.026 (0.104)	-0.129 (0.222)	-0.026 (0.113)
Standardized Composite Language Score	0.251 (0.205)	-0.070 (0.130)	-0.183 (0.209)
Constant	0.452 (0.426)	1.163*** (0.390)	-3.562*** (0.781)
Observations	252	165	127

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

Table 7

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**
	(0.445)	(0.201)	(0.164)	(0.311)	(0.266)	(0.181)
Dummy for inclusion in the Language Study at age 9 months	0.168			0.723*		
	(0.376)			(0.410)		
Constant	9.615	-4.385	-3.486	7.077	4.339	4.315
	(6.647)	(3.399)	(3.641)	(5.657)	(3.778)	(3.233)
Observations	112	108	134	116	102	128
R-squared		0.322			0.366	

Robust standard errors in parentheses, except for the Heckman selection model. We add the following control variables to all of the models shown in Table 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; and dummy for randomization group (block).

*** p<0.01, ** p<0.05, * p<0.1

Table 8

Conversational Turn Counts at Baseline

Variables	Model 1	Model 2	Model 3	Model 4
	Panel A: OLS Estimator			
Random assignment to control or invitation arm	-0.199	-0.116	-0.102	-0.123
	(0.170)	(0.179)	(0.095)	(0.099)
Recording duration			0.471***	0.453***
			(0.043)	(0.042)
Observations	110	108	108	108
R-squared	0.102	0.390	0.519	0.526
	Panel B: 2SLS Estimator			
Attendance to the LENA Start program	-0.397	-0.238	-0.211	-0.249*
	(0.329)	(0.288)	(0.148)	(0.146)
Recording duration			0.451***	0.433***
			(0.028)	(0.030)
Observations	110	108	108	108
R-squared	0.108	0.409	0.526	0.531
Dummies for Randomization Group	Y	Y	Y	Y
Demographic variables	N	Y	Y	Y
Recording duration	N	N	Y	Y
Dummies for Saturday/Sunday	N	N	N	Y

Clustered standard errors in parentheses. We add the following demographic variables to Models 2-4 shown in Table 8: 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 1-4. We add a dummy for recordings that took place on Saturday or on a Sunday in Model 4.

*** p<0.01, ** p<0.05, * p<0.1

Table 9
Impact of the LENA Start Program on Conversational Turns
Lagged Dependent Variable Model

Variables	Model 1	Model 2	Model 3
	Panel A: OLS Estimator		
Random assignment to control or invitation arm	0.130	0.215	0.314**
	(0.187)	(0.160)	(0.113)
Conversational Turn Counts at Baseline			0.609***
			(0.136)
Observations	104	102	90
R-squared	0.452	0.497	0.681
	Panel B: 2SLS Estimator		
Attendance to the LENA Start program	0.232	0.383*	0.551***
	(0.286)	(0.211)	(0.125)
Conversational Turn Counts at Baseline			0.662***
			(0.115)
Observations	104	102	90
R-squared	0.445	0.479	0.680
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

Clustered standard errors in parentheses. We add the following variables to control for differences in recording sessions. First, 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 10
Language Environment at Endline
Fixed Effects Models

Variables	Fixed Effect Estimator	IV Fixed Effect Estimator
Impact of LENA Start (ITT and LATE)	0.323*** (0.072)	0.578*** (0.152)
Observations	214	214
Number of participants	122	122

Clustered standard errors in parentheses. We add the following variables in the regression model. 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 11

Initiation of Conversations in Audio Segments with Focus Child and an Adult

	Panel A: Number of Audio Segments		
	Number of Segments Initiated by the Focus Child	Number of Segments Initiated by a Female Adult	Number of Segments Initiated by a Male Adult
	OLS Estimator		
Random assignment to control or invitation arm	0.394**	0.118	0.248
	(0.155)	(0.178)	(0.177)
Observations	87	87	87
	2SLS Estimator		
Attendance to the LENA Start program	0.690***	0.203	0.432*
	(0.172)	(0.228)	(0.234)
Observations	87	87	87
	Panel B: Number of Conversational Turns by Audio Segment Initiator		
	Conversational Turns in Segments Initiated by the Focus Child	Conversational Turns in Segments Initiated by a Female Adult	Conversational Turns in Segments Initiated by a Male Adult
	OLS Estimator		
Random assignment to control or invitation arm	0.488**	0.238	0.331
	(0.191)	(0.217)	(0.202)
Observations	87	87	87
	2SLS Estimator		
Attendance to the LENA Start program	0.862***	0.414	0.572**
	(0.209)	(0.256)	(0.223)
Observations	87	87	87

Clustered standard errors in parentheses. 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, both for baseline and endline. 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 12
Adult Words in Audio Segments with Key Child and an Adult Person

	Panel A: All Audio Segments with Focus Child and an Adult Person		
	OLS Estimator		
	Adult Words	Female Adult Words	Male Adult Words
Random assignment to control or invitation arm	0.263 (0.169)	0.223 (0.203)	0.336 (0.217)
Observations	87	87	87
	2SLS Estimator		
	Adult Words	Female Adult Words	Male Adult Words
Attendance to the LENA Start program	0.468** (0.200)	0.392 (0.260)	0.587** (0.266)
Observations	87	87	87
	Panel B: Audio Segments Initiated by the Child and Followed by Adult Person		
	OLS Estimator		
	Adult Words	Female Adult Words	Male Adult Words
Random assignment to control or invitation arm	0.495** (0.180)	0.487* (0.220)	0.388** (0.156)
Observations	87	87	87
	2SLS Estimator		
	Adult Words	Female Adult Words	Male Adult Words
Attendance to the LENA Start program	0.886*** (0.220)	0.850*** (0.269)	0.684*** (0.184)
Observations	87	87	87

Clustered standard errors in parentheses. 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 13
Impact of the LENA Start Program on Exposure to TV/Electronics
Lagged Dependent Variable Model

Variables	Model 1	Model 2	Model 3
	Panel A: OLS Estimator		
Random assignment to control or invitation arm	0.046	-0.050	-0.128
	(0.168)	(0.210)	(0.212)
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
	(0.269)	(0.309)	(0.272)
Observations	104	102	90
R-squared	0.199	0.313	0.444

Clustered standard errors in parentheses. We add the following variables to control for differences in recording sessions. First, 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 14
Investigation of Potential Mechanisms of the LENA Start Program

VARIABLES	Parent Self Efficacy	Parent Sense of Social Support	Maternal Beliefs	Maternal Knowledge
	Panel A: OLS Estimator			
Random assignment to control or invitation arm	-0.0305	-0.0770	0.3176*	0.2927
	(0.1576)	(0.1529)	(0.1565)	(0.1846)
Observations	128	128	128	128
R-squared	0.5691	0.3519	0.3183	0.6139
	Panel B: 2SLS Estimator			
Attendance to the LENA Start program	-0.0573	-0.1463	0.6076**	0.5642**
	(0.2583)	(0.2402)	(0.2555)	(0.2643)
Observations	128	128	128	128
R-squared	0.5685	0.3572	0.3124	0.6450

Clustered standard errors in parentheses. 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).

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 1
LENA Start Objective Feedback Report

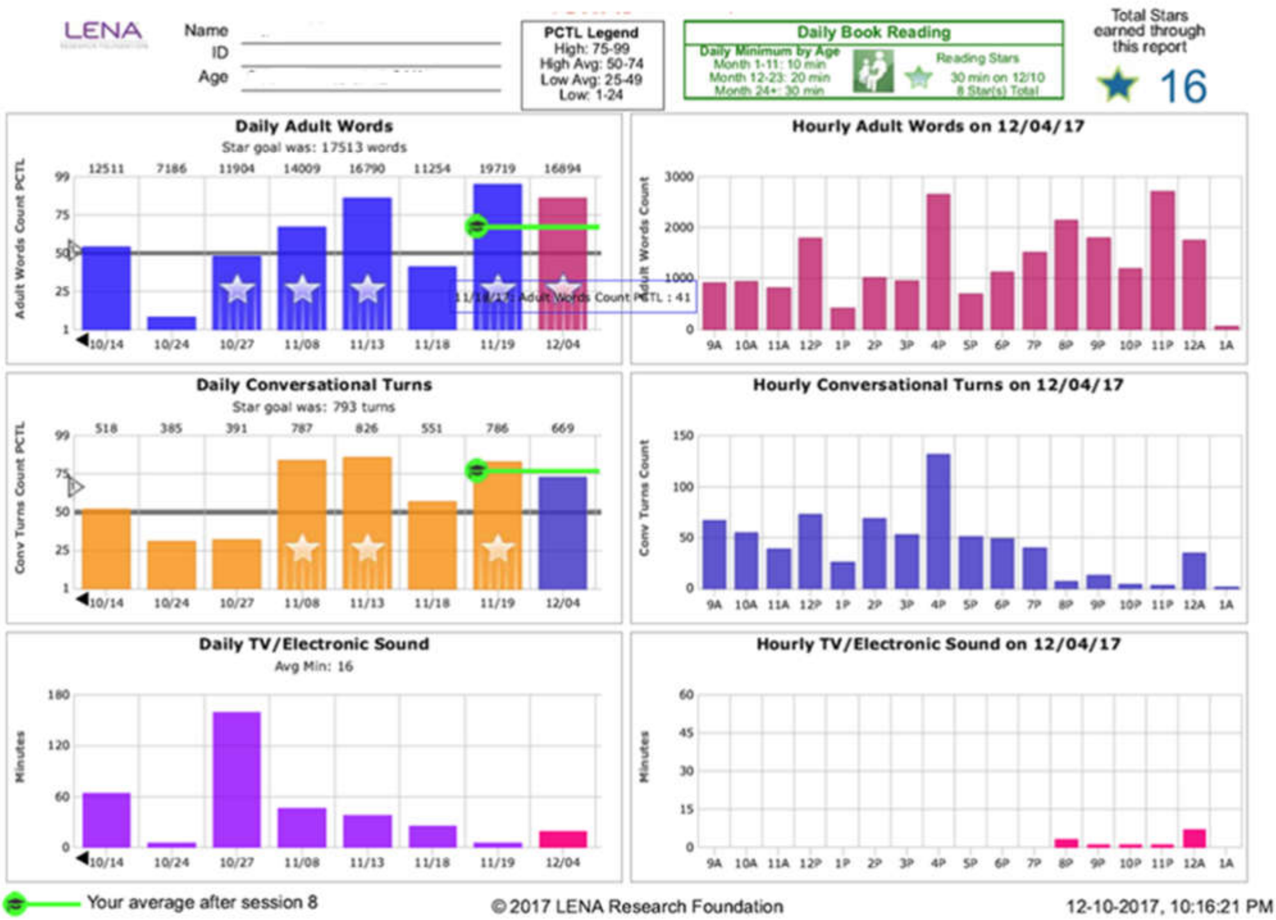
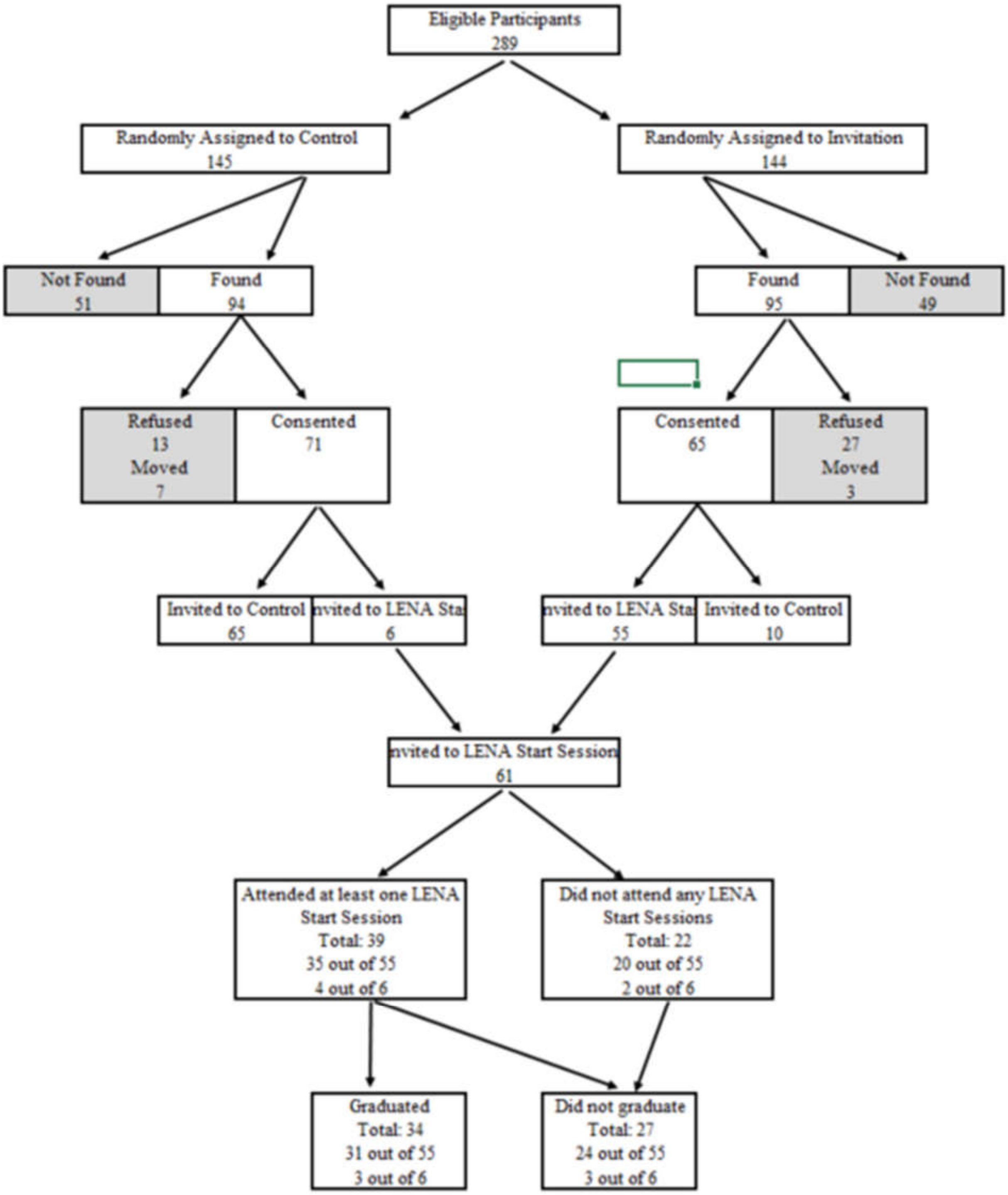


Figure 2

LEME Study: Eligibility, Random Assignment, Recruitment, Consent, Final Assignment, Graduation



Appendix A

A.1 Self-Efficacy and Sense of Support Scales

In this appendix, we describe the operationalization of the data to measure mechanisms. 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 have four items. The four items for 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 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).

To produce scores, we estimate an Item Response Theory Partial Credit Model (PCM). PCM are useful when researchers have ordinal data (as it is in our case). In PCM, each item has its own difficulty parameter and the items share the same discrimination parameter. While it would be possible to allow each item to have its own discrimination parameter, we chose this more parsimonious model because each scale has just four items.

We show the results of our estimates in Appendix Table A1. Appendix Table A2 examines the balance between control and treatment arms at baseline.

A.2 Language Development Knowledge Scale

We reproduce the procedures in Suskind et al. (2015). We score each response with a binary variable (0/1), in which "0" and "1" represent, respectively, a wrong and a right answer (Suskind et al., 2015 provide the answer key). We estimate the average for each topic, and we estimate the overall average. The results in Table 14 use the overall average and the results in Table 14 decompose the analysis by topic (so we use the topic-specific average).

Appendix Table A2 examines the balance between control and treatment arms at baseline. Appendix Table A3 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 smaller impacts on "Learning to Talk" or "Language and TV" subscales. The point estimate of the LATE parameter suggests impacts on "Reading Books," "Learning Math," and the "School Readiness" subscales, but no impacts on the other two scales.

A.3 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 make sure 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 A4, and note that our findings are robust to these choices.

Appendix Table A1

Item Response Theory Analysis

	Panel A: Self Efficacy Scale				Panel B: Social Support Scale			
Item Discrimination Parameter	0.8354***				0.5988***			
	(0.0883)				(0.0730)			
	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	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Alternative 2	1.3009*		0.9799**	1.4050***	1.8367***	0.9687**	-0.2478	0.1594
	(0.7371)		(0.3951)	(0.4223)	(0.6823)	(0.4018)	(0.5421)	(0.4008)
Alternative 3	3.8079***	2.5401***	2.8791***	3.1958***	3.6943***	1.8417***	1.9313***	1.6695***
	(0.6566)	(0.5220)	(0.3900)	(0.4299)	(0.6507)	(0.4167)	(0.4085)	(0.3622)
Alternative 4	5.5949***	3.9296***	3.4348***	3.7778***	4.7321***	2.6067***	2.7363***	1.9392***
	(0.6874)	(0.5455)	(0.4295)	(0.4712)	(0.6650)	(0.4336)	(0.4310)	(0.3981)
Alternative 5	6.5606***	5.2627***	3.5250***	4.1971***	5.2723***	4.0091***	4.0231***	3.5021***
	(0.7048)	(0.5581)	(0.4399)	(0.4805)	(0.6693)	(0.4282)	(0.4307)	(0.3846)
Variance of latent factor	1.0000				1.0000			
	(0.0000)				(0.0000)			
Number of observations	266	266	266	266	266	266	266	266

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix Table A2

Relationship Between Baseline Data for Mechanisms and Treatment Assignment

VARIABLES	Parent Self Efficacy	Parent Sense of Social Support	Maternal Beliefs	Maternal Knowledge
Random assignment to control or invitation arm	0.1981	0.0475	0.2911	0.1884
	(0.1575)	(0.1911)	(0.2058)	(0.1153)
Constant	3.4149	4.6713	-3.5165	3.9236*
	(3.2691)	(5.0799)	(3.4895)	(2.1016)
Observations	134	134	134	134
R-squared	0.2113	0.1144	0.1004	0.4575

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

Table A3

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	0.1174	0.2225*	0.3586	0.2500	0.1890
	(0.1833)	(0.1117)	(0.2147)	(0.1564)	(0.1706)
Observations	128	128	128	128	128
	Panel B: 2SLS Estimator				
Attendance to the LENA Start program	0.2229	0.4277***	0.6850*	0.4784*	0.3619
	(0.2821)	(0.1420)	(0.3711)	(0.2557)	(0.2613)
Observations	128	128	128	128	128

Clustered standard errors in parentheses. 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

Appendix Table A4
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	0.3176*	0.2646*
	(0.1565)	(0.1343)
Observations	128	128
	Panel B: 2SLS Estimator	
	0.6076**	0.5051**
Attendance to the LENA Start program	(0.2555)	(0.2151)
Observations	128	128

Clustered standard errors in parentheses. 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

Appendix B

In Appendix B, we report our findings about the sensitivity of estimates of the program's impact to attrition due to valid recording (Table B1) and the specification of recording duration in the regression equation (Table B2).

Table B1

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

	Inverse Probability Weight	Heckman Selection	Lee Sharp Bounds
Random assignment to LENA Start Program	0.295* (0.150)	0.314** (0.149)	
Lower Bound (Lee Sharp Bounds)			0.091 (0.148)
Upper Bound (Lee Sharp Bounds)			0.330** (0.152)

Robust standard errors in parentheses, except for the Heckman Selection model. We add the following control variables to all of the models shown in Table 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; and dummy for randomization group (block).

*** p<0.01, ** p<0.05, * p<0.1

Appendix Table B2

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

	Fixed Effect Specification			
	(1)	(2)	(3)	(4)
VARIABLES	Linear	Quadratic	Cubic	Quartic
	Panel A: OLS Estimator			
Random assignment to control or invitation arm	0.323***	0.326***	0.338***	0.323***
	(0.072)	(0.063)	(0.063)	(0.075)
	Panel B: 2SLS Estimator			
Attendance to the LENA Start program	0.578***	0.582***	0.600***	0.580***
	(0.152)	(0.142)	(0.149)	(0.167)

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