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

Flávio Cunha[†], Marsha Gerdes[‡], Qinyou Hu[§], and Snejana Nihtianova[**]

January 1, 2023

Abstract

Research documents that parental beliefs influence early investments in children, which, in turn, determine early human capital and, eventually, other skills children acquire in later stages of the lifecycle, such as literacy. Our paper reports the results of an experimental evaluation of the LENA Start Program, a group- and center-based parenting program that teaches the science of early language development, models verbal interaction behaviors with children, and provides objective feedback to improve the early language environment. The intervention changes parental beliefs and impacts the quantity and quality of parental linguistic input.

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

[*]We thank Delitza Hernandez, Michelle Gifford, Shanae Smith, and Mary Triguba for their 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 supported 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 national sites, energized our team. We are grateful to Jeffrey Richards, who generously shared crucial knowledge on operationalizing the LENA system data 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. Finally, 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, the 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 and NBER.
[‡] Children's Hospital of Philadelphia.
[§] Rice University.
[**] Rice University.

I. Introduction

The 2017 National Assessment of Educational Progress 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 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 (e.g., Hanushek et al., 2015; Murnane et al., 2000; Vignoles et al., 2011). Beyond financial returns, research also shows that the ability to comprehend written text is associated with better access to health services (e.g., Dexter et al., 1998; Sandiford et al., 1995), improvements in self-esteem (Bown, 1990; Canieso-Doronila, 1996), higher levels of empowerment (Burchfield et al., 2002; Kagitcibasi et al., 2005) and political participation (Egbo, 2000; Hannum & Buchmann, 2003; Ireland, 1994; Purcell-Gates & Waterman, 2000), promotion of cultural change, and cultural diversity preservation (Carr-Hill et al., 2001; Chebanne & Nyati-Ramahobo, 2003; Norwood, 2003; Robinson-Pant, 2000; Tarawa, 2003). Reading proficiency is so fundamental in modern life that it is considered a basic human right (UNESCO, 1975).

Theoretically, reading proficiency requires two distinct skills: decoding and language comprehension skills (Gough and Turner, 1986; Scarborough et al., 2009).[2] Empirically, most children who cannot read proficiently fail to do so because they lack appropriate language skills (Foorman, Petscher, and Herrera, 2018, OECD, 2016). In addition, a large body of research has documented steep socioeconomic (SES) gradients in language development observed before the school years (Fernald et al., 2012; Hart & Risley, 1995; Hoff, 2006; Huttenlocher et al., 2010).

Developing these critical language skills requires greater exposure to language so that children can practice and hone their language processing skills and simultaneously increase their vocabulary (e.g., Weisleder and Fernald, 2013; Pan et al., 2005; Gilkerson et al., 2018; Romeo et al., 2018). This literature finds 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 (e.g., Hart and Risley, 1995; Hoff, 2003, 2006; Rowe and Goldin-Meadow, 2009; Fernald et al., 2012).

[1] Low-income children are children eligible for the National School Lunch Program.
[2] We present a detailed literature review in Appendix A.

Recent literature shows that low-income mothers underestimate the returns on early investments (e.g., Cunha et al., 2013; Attanasio et al., 2019). In addition, heterogeneity in parental beliefs predicts early parental investments (e.g., Attanasio et al., 2019; Cunha et al., 2022; Biroli et al., 2022). Indeed, Bhalotra et al. (2020) show that once they control for perceived returns and costs of early investments (i.e., breastfeeding), there is little role for heterogeneity in preferences. In addition, this research also finds sizeable socioeconomic gradients in parental beliefs (Boneva and Rauh, 2018; List et al., 2021).

Low expectations about the returns to investments might rationalize the existence of parenting interventions in the developed (e.g., the Nurse-Family Partnership in the USA, see Olds et al., 2004; Heckman et al., 2017) and developing worlds (e.g., the Jamaica Home Visiting Program, Grantham-McGregor et al., 1991; and its replications in other countries such as the ones described in Attanasio et al., 2014; Carneiro et al., 2019; and Heckman et al., 2020). In parallel, developmental psychologists have created parent-directed interventions to improve early language environments (Leech et al., 2018, McGillion et al., 2017a; Suskind et al., 2015; Rowe and Leech, 2019; and Ramirez et al., 2020). These studies show that interventions that combine education (i.e., teaching parents about the science of brain and language development), coaching (i.e., modeling language interaction behaviors to parents), and objective feedback (i.e., providing objective statistics about the quality of the language environment) might produce impressive impacts on language outcomes. These interventions might work by elevating parental expectations about the returns to early investments.

This paper reports the results of evaluating a scalable intervention designed to improve disadvantaged children's language environment, known as the LENA Start Program. The LENA Start Program occurred within the Philadelphia Human Development longitudinal study (henceforth, PHD Study). This study recruited 822 mothers during the second trimester of their first pregnancy and elicited mothers' subjective beliefs about the importance of parent-child interaction for child development. When the children were nine to twelve months old, the study visited the families in their homes and measured family investments. In addition, the study assessed cognitive and language development when the children were two years old.

In the PHD Study data, the language environment's heterogeneity – measured by conversational turn counts – has a fragile association with family income. Our research hypothesizes that heterogeneity in parental beliefs drives parental linguistic input gaps. Therefore, 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.

The LENA Start Program is ideal for testing our hypothesis because of its curriculum. It explains 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 typical parenting programs, LENA Start is a low-cost group-centered model that may help parents build social capital and a solid social support network.[3,4]

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. We assigned the other half to receive an invitation to participate in the LENA Start Program. We successfully located 94 (or 65%) of the 145 parents in the control group and 95 (or 66%) of the 144 parents in the invitation group. Once we found the parents, we tried to consent the parents to participate in the study activities. Nearly 75% (71 out of 94) of parents in the control group agreed to participate in the study. Approximately 68% (65 out of 95) of the treatment-group parents decided to participate in the study. Therefore, we finally enrolled 136 families in the LENA Start evaluation study. Due to the group format, our intervention suffers from an issue of non-random attrition. We admit that non-random attrition can compromise random assignment. However, Section II. A. and Appendix C show that our sample is balanced conditional on the random assignment and that the random assignment indicator strongly predicts the final attendance status. Appendix F shows that the treatment modality does not drive our sample attrition. The results are robust to attrition correction using inverse probability weighting Wooldridge (2010) or the Lee (2009) treatment effects bounds.

The evaluation study's primary outcomes are objective measures of the language environment, which we obtained with the LENA System (Gilkerson and Richards, 2008). 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. To complement the LENA System, one of the innovations of our measures is that we also use 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). The LENA System estimate does not distinguish between words directed

[3] Zhang et al. (2015) and Gilkerson et al. (2018) are interventions that use linguistic feedback to parents.
[4] As of Spring 2019, the LENA Start Program is being implemented by hospital systems, school districts, public libraries, and NGOs that serve parents and their young children. For a list of current LENA Start program sites, see https://www.lena.org/about/#where-is-lena.

to the child and words overheard by the child. However, with LENA ADEX, we can identify who initiated the conversation (the child or a male or a female adult) and the number of conversational turns per segment as categorized by the segment initiator (the child, a female adult, or a male adult). In that way, we can investigate whether children or adults initiate more conversations (audio segments).

We also collected survey measures at both baseline end endline on four potential mechanisms that could explain any positive findings of the impact of the LENA Start Program on parent-child linguistic interaction. Specifically, we measured (1) parental beliefs and (2) parental knowledge about the importance of the language environment for a child's language development. We also measured (3) parental self-efficacy, defined as the parent's belief in their ability to perform the parenting role competently, and (4) their sense of social support perception. We followed the instrument design of Cunha et al. (2013) to measure parental beliefs. First, we presented mothers with two hypothetical scenarios of language environment ("investments") and asked them to predict the child's future language development. Next, we used the instrument Suskind et al. (2016) developed to measure parental knowledge. Then, we followed Bandura (1977) to measure parental self-efficacy. Finally, we used four items to measure their social support perception due to the group nature of this particular intervention.

We analyze the program's Intent to Treat (ITT) and Local Average Treatment effects (LATE). We find strong evidence that the LENA Start Program significantly impacts conversational turns and adult words in conversations between the focus child and an adult. Controlling for conversational turns at baseline, we find the LENA Start Program significantly increases conversational turns between adults and the focus child by 31.4% of a standard deviation. With the measures from LENA ADEX, we find that the differences in conversational turns arise because the LENA Start Program children initiate a more significant number of conversations – ITT estimates indicate that the treatment group children begin nearly 40% of a standard deviation more conversational turns than their control counterparts. The Program also significantly increases adult words spoken to the child. The significant increase is the product of a higher number of conversational turns initiated by children and responded to by parents in ways that do not stop the conversation but encourage the children to continue to talk and interact. Finally, we find the Program does not impact exposure to TV or other electronics.

Exploring the mechanisms, we find that the change in parental beliefs about the language environment's importance for language development explains the program's impacts. Indeed, our mediation analysis shows that changing parental beliefs can explain at least 34 percent and up to 54 percent of the total program impact. 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. (2015), 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 discussed below, these ingredients are present in the LENA Start Program. Our innovation is that we evaluate not a prototypical program but a program that school districts, public libraries, and children's hospitals have adopted nationally. In fact, according to the LENA Foundation, nearly 2,000 families participate in the LENA Start Program annually.

Our paper also contributes to the literature to the growing literature that investigates the malleability of parental beliefs. For example, Attanasio et al. (2019) found that a home-visiting program in Colombia did not shift parental beliefs, which is consistent with the result that the same intervention did not produce permanent changes in family investments. In contrast, Carneiro et al. (2019) evaluated a large-scale, low-cost parenting program and found that the intervention raised beliefs and improved children's early environments. Finally, List et al. (2021) showed that an intervention combining education and feedback delivered through home visits increases beliefs and parent-child interactions. Our work contributes to this literature by showing that low-cost, center- and group-based interventions that combine education, coaching, and objective feedback improve children's early environments through the malleability of parental beliefs.

We organize the rest of this paper in the following way. Section II describes our study procedures and details of the LENA Start Program, including participants, recruitment, intervention, and measurement. We also present the empirical methodology for evaluating the program's effects. Section III 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 IV concludes.

II. Study Procedures

This section describes the study procedures approved by the University of Pennsylvania and Children's Hospital of Philadelphia Institutional Review Boards.

A. Participants

We used three criteria to determine LENA Start Program Evaluation Study eligibility. First, we restricted eligibility to PHD Study mothers who resided in the inner-city and whose children were at most 33 months by May/2017 because the LENA Start Program is designed for children up to that age. [5] Additionally, we attempted to recruit mothers who authorized the research team to be contacted for participation in future PHD Studies. We identified a group of 289 mothers who satisfied all of the three inclusion criteria.

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." Figure 1 describes the results of the randomization procedure. We randomly assigned 145 of the 289 eligible mothers to the control group and the remaining 144 to the group invited to participate in the LENA Start Program.

The group format of the LENA Start Program creates specific challenges for the design of the evaluation protocol. For example, the low-income families participating 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. On the other hand, the longer the recruitment period, the more likely parents who were recruited early would report a change in the work schedule and could no longer attend group sessions when they had agreed to participate in recruitment (to the LENA Start Program). Therefore, our team decided to make an intense effort to recruit participants within three weeks and start the Program in the

[5] Appendix B contains additional information about the PHD Study.

fourth week. 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.

Next, we describe the protocol for the recruitment of participants. Consider a block with twenty-six mothers, ordered in descending order of the draw from the uniform random variable. Therefore, the mother ranked first is the one who drew the largest realization of the random draw, while the mother with the smallest draw is ranked last. We invited the top thirteen mothers to participate in the LENA Start Program. In addition, a few of the bottom thirteen may receive an invitation to join if we need additional participants to form a LENA Start group.[6]

We invited the mothers at the bottom half of the ordered invitation list to participate in the study's control arm. First, we called the control mothers from bottom to top. 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 will explain in the next section.

We invited the mothers to participate in the study's treatment arm at the top of the ordered invitation list from top to bottom (i.e., the opposite for the control group). 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 join the LENA Start Program, our research assistants invited that mother to participate as a control group member. We did so for two reasons. First, we had little time to recruit parents for the evaluation study. Because of our families' unstable living arrangements, we knew 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 expectations from the literature that reports the demand for parenting education programs (e.g., Kalil, 2014). Studies that follow the typical protocol do not usually collect data from parents who refuse to participate as treatment group members. As a result, we typically know very little about such parents. Because they join as members of the control group (although no longer in a randomized fashion), we can learn more about these

[6] We invited six control mothers to participate in the LENA Start Program to ensure we reached a minimum number of mothers in some groups.

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.

As shown in Figure 1, once we found the parents, we tried to consent the parents to participate in the study activities. As a result, nearly 75% (71 out of 94) of the parents in the control group agreed to participate in the study. On the other hand, approximately 68% (65 out of 95) of the treatment-group parents decided to participate in the study.

As we explained above, accepting the LENA Start Program was not random because some families randomly assigned to receive the invitation declined the offer. Therefore, our team invited a few control group members to have groups with ten families. As shown in Figure 1, the final acceptance list contained 61 families, 55 of which came from the study's invitation arm, and six 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 Table 1, we show the results of the balance check. First, in Panel A, we compare the PHD Study participants eligible for participation in the LENA Start Evaluation Study with non-eligible ones. We find 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%. We note that the LENA measures indicate a lower quality of language environment because the conversational turns are higher and exposure to TV is lower for the control group children (even though the differences are not statistically significant). Second, we look at the post-attrition sample in Panel B. There are no differences between the two groups except for the mother's marital status and HOME scores. In the post-attrition sample, those in the treatment group are more likely to be single and unmarried. Appendix C provides more details about the balance and attrition of our study sample.

B. Intervention

The LENA Start Program aims to improve parental linguistic input by improving parental knowledge about the importance of the early language environment for language development and providing tips for enriching the early language environment.[7] The Program lasts thirteen weeks, and during this period, groups of ten to twenty families meet for about one hour a week

[7] The website https://www.lena.org/lena-start/ provides an overview of the LENA Start Program.

to build family engagement and social capital. The program has five components: Education, Coaching, Feedback, Book Reading, and Language Development Reporting.

Education: The program provides information about the importance of the language environment for a child's language development. Rowe and Goldin-Meadow (2009) show that 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 documenting infants' capacity to engage in social interaction and that parents and infants routinely use nonverbal communication. For example, in one of the sessions, the program presents the "Still Face Experiment" by Tronick et al. (1978). The experiment involves an infant and a 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 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.

Coaching: Each session leverages "Talking Tips" 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. Parents are encouraged to share successful strategies when they return to a later session. They use their own experience to generate new talking tips that other parents in the group can also 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, what interests them, conversations, and 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 infants are more likely to engage in joint attention when the parent refers to an object 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 the 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., 2015). 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.

Feedback: 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 have the chance to review their information 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 they can and, at the same time, point out areas where 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 hourly (for the previous recording day).

Figure 2 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 offer 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 previous recording day. The green text on the top right corner of the feedback report contains a parent self-report number of reading minutes. In addition, the feedback report includes the number of stars parents have received to date. The parent gets 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 their best.

Next, the Coordinator uses the report to identify opportunities for improvement. For example, when reviewing the information, the Coordinator will remark that the parent and the child interacted a lot (as measured by conversational turns by the hour, as reported in the middle row and right column of Figure 2) 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, not the one passing judgment on a parent's skills. The primary focus is on providing feedback about conversational turns. 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 children's books. Each session incorporates tips on reading books, talking about books to children, and encouraging parents to read to their children. 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., 2018). The Snapshot is an instrument designed to evaluate children's language skills. 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 screen for signals of severe language development delays that may require additional attention from an early intervention specialist.

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. A parent graduates from the LENA Start Program if the parent satisfies two requirements. First, the parent attends all of the first four sessions. Second, the parent had to complete at least five of the remaining nine sessions.[8]

C. Measurement

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 were identical for control and treatment groups and the same before (baseline) and after the program (endline).

We describe the logistics of the recording activities carried out by families participating in the study. The logistics described in this section relate to the data collection to evaluate the LENA Start Program. As such, we describe a protocol applied to families in the study's control and treatment arms. Each parent received a LENA recorder and a child's vest that had a front pocket for the recorder. Parents recorded the child's audio environment once at baseline and once at the endline. We asked the parent to let the device run for at least 12 hours to record the child's audio environment. In addition, 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.[9] 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 we report below.

The LENA System identifies and labels individual adult and child utterances (also called segments), among other sounds. One conversational turn is a speech-related vocal utterance initiated by the focus child (or adult) with a response 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

[8] If the parent could not attend a session, they could contact the LENA Start Coordinator to schedule a make-up session.
[9] The estimates of the language environment variables LENA System produces are highly accurate (see Gilkerson and Richards, 2008).

we use in our paper because it correlates with adult-child joint attention and speech recasting.[10] Golinkoff et al. (2019) argue that "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."

The LENA System estimates the number of adult words spoken by adults 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 & 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 until children are about four or five years old (Messenger et al., 2015; Rollins, 2003).[11] As explained by Golinkoff et al. (2019), overheard speech does not support a rich 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, our analysis explores 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.

One of the innovations of this paper is that we have an indirect way of assessing whether the improvements in conversational turns are "just for show" or reflect fundamental 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). In so doing, we can then identify who initiated the conversation (the child or a male or a female adult) and the number of conversational turns per segment as categorized by the segment initiator (the child, a female

---

[10] See Goldin-Meadow et al. (2014); Harris et al. (2010); Malin et al. (2014); McGillion et al. (2017); McGillion et al. (2017b); 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.

[11] Studies of overheard speech may overestimate its impact on language development because they typically eliminate or at least significantly reduce other stimuli in the lab. Hence, they minimize attentional demands from children. See, for example, Yuan and Fisher (2009).

adult, or a male adult). Appendix D provides a detailed description of our procedures to clean and create the final dataset for analysis.

We interviewed the parents who agreed to participate in the study to understand the mechanisms of the intervention. The baseline and endline surveys occurred just before and after the LENA Start Program, respectively. 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. The baseline and endline survey questionnaire had only eleven questions and focused on measuring four potential mechanisms.

We measure four potential mechanisms. First, our survey questionnaire also measures parental self-efficacy (Bandura, 1977), 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) alternative that ranges from "least sure" to "very sure." For example, one item states, "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 the parental self-efficacy scores predict the child's psychological functioning and adjustment. They also found that parents with higher self-efficacy scores had higher parenting competence and satisfaction levels (Leahy-Warren et al., 2012).

Second, 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) alternative that describes 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 et al., 2012).

We estimate an Item Response Theory Partial Credit Model (IRT-PCM) to produce self-efficacy scores and sense of support scales. IRT-PCM is useful when researchers have ordinal data. While allowing each item to have its discrimination parameter would be possible, we chose this more parsimonious model because each scale has just four items. We provide additional detail about these three measures in Appendix E.

Third, we measured parental knowledge of language development. We used the instrument Suskind et al. (2016) 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.[12] The second subscale measures parental attitudes about reading to children.[13] The third subscale assesses parental perceptions about how children learn math.[14] The fourth subscale evaluates parental perceptions about the relationship between language development and school readiness.[15] The fifth and last subscale quantifies parental perceptions about TV and language development.[16] For each item, parents choose one alternative (among five) that describes the extent to which they agree with the statement. The alternative ranges from "Strongly Disagree" to "Strongly Agree." We estimate the global and topic-specific scores.

Fourth, we adapted the procedure in Cunha et al. (2013) to elicit parental beliefs about the importance of the language environment for language development. The elicitation instrument has two items reflecting low and high scenarios of investments and asking 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 environment, with lots of exposure to TV and little verbal exchange between adults and children. Finally, we ask parents to predict the child's language development at three years old. Parents choose one of the following five alternatives: "Low," "Low Average," "Average," "High Average," and "High," and we assign these alternatives to percentiles 5, 25, 50, 75, and 95, respectively.

We follow Cunha et al. (2022) in constructing parental beliefs. First, we assume that language development at age three follows a normal distribution with mean zero and variance one. Then, we compute the scores of the standardized normal for each one of the percentiles. Let $s_{i,1}$ and $s_{i,2}$ denote mother i's score in the first and second scenarios, respectively. The difference $\Delta_i = s_{i,1} - s_{i,2}$ represents our error-ridden measure of parental beliefs.

Cunha, Elo, and Culhane (2022) show that the estimated parental beliefs contain much measurement error. Therefore, we combine the error-ridden measure above with three items

---

[12] The second item in this subscale states that: "Children learn fewer words when adults talk with a warm tone."
[13] The first item in this subscale states that "You cannot teach children anything new by reading them the same book over and over."
[14] The third item in this subscale states that "Talking about the difference between tall and short teaches toddlers about math.
[15] 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."
[16] The last item in this subscale states that "The more television children under two watch by themselves the more words they learn."

from the Suskind et al. (2016) knowledge scale.[17] We factor analyze the four measures and predict the factor score using the Bartlett formula. This factor score is our error-adjusted estimate of parental beliefs.

D. Methods

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 (LATE)," which uses Z_i as an instrumental variable for D_i.

Let $Y_{i,t}$ denote child i's conversational turns at period t. The variable $B_{i,j}$ is equal to one if mother i is in randomization block j, and zero, otherwise. Let $R_{i,t}$ denote the vector that contains information about child i's LENA recording at period t, where $t = 0$ is the baseline, and $t = 1$ is the endline. This vector includes the recording duration and dummies for a recording that occurred on Saturday or Sunday. Finally, the variable X_i captures demographic characteristics. We estimate the following regression model to obtain the ITT:

$$Y_{i,1} = \beta_0 + \beta_1 Z_i + \beta_2 Y_{i,0} + X_i \beta_3 + \sum_{t=0}^{1} \beta_{4+t} R_{i,t} + \sum_{j=1}^{J} \gamma_j B_{i,j} + \epsilon_{i,1}. \quad (1)$$

We now turn to the analysis of the impacts of the LENA Start Program on parents who attended at least one session. We estimate the LATE from the following two-stage least squares regression model:

$$D_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 Y_{i,0} + X_i \alpha_3 + \sum_{t=0}^{1} \alpha_{4+t} R_{i,t} + \sum_{j=1}^{J} \pi_j B_{i,j} + \omega_{i,1}. \quad (2)$$

$$Y_{i,1} = \beta_0 + \beta_1 D_i + \beta_2 Y_{i,0} + X_i \beta_3 + \sum_{t=0}^{1} \beta_{4+t} R_{i,t} + \sum_{j=1}^{J} \gamma_j B_{i,j} + \epsilon_{i,1}. \quad (3)$$

[17] We selected the three items with the strongest correlation with our error-ridden measure of parental beliefs. They are items #22 ("Talking to children cannot make them smarter"), #23 ("How many words three-year-olds know can predict how well they might do in kindergarten"), and #30 ("The more television children under two watch by themselves the more words they learn").

For both ITT and LATE, we consider three variations of the models. The first model imposes $\beta_2 = \beta_3 = \beta_4 = 0$. Thus, it does not include the lagged outcome, covariates for demographic characteristics, or baseline recording information (*Model 1)*. The subsequent models have each one of these components, one at a time. In the second model, we include demographic characteristics (so we relax the restriction that $\beta_3 = 0$) (*Model 2)*. Finally, we add baseline conversational turn counts and recording information and estimate all the coefficients in (1) (*Model 3)*.

The error term $\epsilon_{i,1}$ 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 experiences and exchange ideas about implementing the "Talking Tips." However, due to the insufficient number of clustered groups (we only have ten groups), we also provide randomization inference p-values. We use the Young (2018) permutation test with 2,000 stratified clustered resamplings. We also report the conventional robust standard errors for all the regression results for comparison purposes. We show the robustness of our point estimates and the corresponding methods in Appendix F.

Let $M_{i,k,t}$ denote child i's mediator k at period t. We analyze two regression equations in the mediation analysis. The first equation estimates the impact of the LENA Start Program on each one of the mediators separately:

$$M_{i,k,1} = \delta_{k,0} + \delta_{k,1} Z_i + \delta_{k,2} M_{i,k,0} + X_i \delta_{k,3} + \sum_{j=1}^{J} \theta_{k,j} G_{i,j} + \eta_{i,1}. \quad (4)$$

The parameter of interest is $\delta_{k,1}$ in equation (4). The second equation expands equation (1) by including mediator k:

$$Y_{i,1} = \tau_{k,0} + \tau_{k,1} Z_i + \rho_k M_{i,k,1} + \tau_{k,2} Y_{i,0} + X_i \tau_{k,3} + \sum_{t=0}^{1} \tau_{k,t+4} R_{i,t} + \sum_{j=1}^{J} \phi_{k,j} G_{i,j} + \zeta_{i,k,1}. \quad (5)$$

In this equation, we are interested in the parameters $\tau_{k,1}$ and ρ_k. In this mediation exercise, the parameter β_1 in (1) captures the LENA Start Program's total effect on conversational turns. We can decompose the total effect through the formula $\beta_1 = \tau_{k,1} + \rho_k \delta_{k,1}$, where $\tau_{k,1}$ and $\rho_k \delta_{k,1}$ are the direct and indirect effects, respectively. As elsewhere in our analyses, we estimate and present robust and clustered standard errors.

III. Results

A. Conversational Turns

In this section, we present the program's impacts on conversational turns because it is our primary variable of interest. Table 2 reports our findings regarding the impact of the LENA Start Program on conversational turn counts. Panel A presents the OLS estimators of β_1 and β_2 of variations of (1). 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 baseline conversational turns and recording characteristics, the ITT's point estimate is 31.4% of a standard deviation and statistically significant at a 5% confidence level.

Panel B in Table 2 presents the results for LATE. According to Model 1, the program's impact on conversational turns, 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 is statistically significant at the 10% level. If we hold the conversational turn counts at baseline constant, the LATE increases to 55.1% of a standard deviation and is statistically significant at the 1% level.

These large impacts may arise because parents in the treatment group converse much more with the children but only when the children wear a LENA recorder. It would be challenging to design and implement an ethical study that comprehensively addresses this concern as any such research would have to disclose all of the study procedures (including recording the child's language environment). We thus use the ADEX audio segmentation tool to investigate whether children or adults initiate more conversations (audio segments). If the adults change their behavior because of the recorder, then the differences in conversational turns between the control and treatment groups will reflect the differences in conversations initiated by the 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 respond in ways that require a response from the child. Therefore, 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 a conversation initiated by the focus child. This behavior change is consistent with the LENA Start curriculum.

For this analysis, we use the lagged variable model described in equation (1) to estimate the ITT parameter and the model described in equations (2) and (3) to compute the LATE parameter. Panel A in Table 3 presents the results of our analysis of the initiation of segments of conversation. The OLS and the 2SLS estimators show that the differences in conversational turns arise because the LENA Start Program children initiate more conversations. According to the ITT, the treatment group children begin nearly 40% of a standard deviation more conversations than their control counterparts. The LATE parameters suggest that the impact is 69% of a standard deviation. In contrast, the estimates are statistically significant for conversations initiated by female or male 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. In addition, 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, this finding suggests spillover within families.

In summary, the LENA Start parents do not initiate more conversations than the control group's parents. However, LENA Start children initiate more conversations, and when they do so, the LENA Start parents respond in ways that lead to more conversational turns. These findings reduce (but do not eliminate) the concern that the increase in conversational turns occurs when they wear the device.

These results 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 talk about objects of interest to the child, and thus the conversations occur during joint attention events.

B. Robustness

Appendix F extends our analyses to include three critical robustness checks. First, we used lagged dependent variables in equations (1), (2), and (3) to account for the fact that the families in the control group had higher counts of conversational turns at baseline (see Table D2). Appendix F shows that the point estimates are similar if we use a fixed-effect specification. For example, the ITT is .312 in the lagged dependent model but .323 in the fixed-effect specification. However, the standard errors are larger, and the impacts are no longer significant at the 5% level.

Second, we investigate how attrition impacted our findings (see Table F2). The ITT reduces to .295 (from .312, a 5% reduction) if we use Inverse Probability Weight to reweigh observations. However, it is .314 when we estimate a Heckman Selection model. The Lee (2009) bounds indicate that the ITT is between .056 (with a standard error of .160) and .366 (with a standard error of .158). The lower bound is not statistically significant at the 5% level, but the upper bound is.

Finally, recording duration enters linearly in equations (1) and (3). Therefore, we considered specifications with polynomials of degrees two to four (see Table F3). When we do so, we find that the ITT and LATE estimates are slightly larger then the linear specification we report in Table 2.

C. Other Measures of the Quality of the Language Environment

The LENA System produces two additional measures of the quality of the language environment: adult words and the amount of time of sounds from TV or other electronics. Unfortunately, the raw measures are very noisy. For example, adult words may capture words in a telephone conversation when the child is nearby. As we explain in Appendix A, this speech contributes little to early language development.

We use the ADEX to reduce the noise in these measures (see Table G1). First, we restrict adult words in audio segments where there is at least a conversational turn between the child and an adult. When we do so, we find that the ITT is .263 but not statistically significant, while the LATE is .468 and significant at the 5% level.

Second, we restrict adult words in audio segments initiated by the focus child. These segments drive the impact on conversational turns (see Table 3). When we do so, we find that the ITT and LATE are .495 and .886, respectively. Also, they are significant at the 5% and 1% levels, respectively. This evidence reinforces the findings on conversational turns.

In contrast, we do not find any impact of the program on TV or other electronics. The point estimates are small (see Table G2).

D. Mediation Analysis

Finally, we study the potential mechanisms of the impacts of the program. As described in Section II.C., we measured four possible mechanisms parental self-efficacy, parental sense of social support, parental knowledge, and parental beliefs. Appendix C contains detailed information about how we operationalize these constructs.

Table 4 presents our estimates for the parameter $\delta_{k,1}$ in equation (4). The ITTs for parental self-efficacy and parental sense of social support are close to zero and not statistically significant. In contrast, the ITT for parental knowledge is .293, and borderline statistically significant. The ITTs for parental beliefs are .306 and .451 for the error-ridden and factor score measures, respectively, and are statistically significant at the 10% and 5% levels, respectively. The LATE parameters also rule out any impacts of the LENA Start Program on parental self-efficacy or sense of social support. In addition, it confirms that the program impacted parental knowledge and parental beliefs.

Table 5 presents our estimates of the parameters ρ_k in equation (5). The ITT reduces by 20% when we use the error-ridden measure of parental beliefs, but it is still statistically significant, while the parameter ρ_k is 0.098 and not statistically significant. However, the ITT reduces by nearly half when we add the factor score for parental beliefs, which is no longer significant. In this case, ρ_k is 0.198 and significant at the 5% level.

We execute a Monte Carlo simulation to estimate the distribution of the indirect effect $\rho_k \delta_{k,1}$. For this analysis, we focus on the factor score as the relevant measure. Figure 3 presents our findings for the models in which we estimate clustered (top) or robust (bottom) standard errors. We estimate that the indirect effect is 13.7% of the total effect. The 95% confidence intervals are [0.026, 0.288] and [0.002, 0.321] for the models with clustered and robust standard errors. Thus, we conclude that parental beliefs partially mediate the effects of the intervention.

In contrast, table G3 presents the results for the other three mediators. We cannot reject the null that the coefficient ρ_k is zero, which suggests that the other three variables do not mediate the intervention's impact on the quality of the language environment.

IV. Conclusion

The LENA Start Program improves children's language environment. These findings matter because parental linguistic input correlates with early language processing skills, vocabulary growth, higher language development in early adolescence, and more robust activation of Broca's brain area, which, in turn, are linked to literacy development. In addition, differences in parental response to conversations initiated by the child drive the program's impacts.

We tested four different mechanisms. First, we found no evidence in favor of parental self-efficacy, parental social support, or parental knowledge. Instead, we found evidence that the program's impact is due to improvements in parental beliefs about the importance of a rich early language environment for a child's language development. Our findings indicate that parental beliefs are malleable and that changes in parental beliefs can change investment behaviors. This result is consistent with the findings of the evaluation of the Nadia Es Perfecto – Intensive (NES-I, Carneiro et al., 2019) and the 3T Home Visitation Program (List et al., 2021). These interventions feature unstructured curricula that include education, coaching, and feedback.

In contrast, Attanasio et al. (2019) concluded that Colombia's Reach Up Program did not change beliefs. There are at least three explanations for this difference: Curriculum, target population, and implementation fidelity. In common, the Reach Up and LENA Start interventions provide education and coaching. In contrast, Reach Up contains a highly structured curriculum with much less opportunity for feedback. Therefore, one possible explanation is that lack of structure or objective feedback is crucial for impacting parental beliefs.

Another possible explanation is the target population. For example, research shows that the original implementation of the Jamaica Home Visiting Program led to significant short-term effects on language and cognition (Grantham-McGregor et al., 1991) and long-term effects on relevant socio-economic outcomes (Gertler et al., 2014; Gertler et al., 2021). Attanasio et al. (2022) explain that the intervention in Jamaica targeted undernourished or severely undernourished children. Interestingly, in Colombia, the program targeted low-income families who participated in a Conditional Cash Transfer program, and there was no indication that the children in these families were undernourished. This explanation suggests that the Reach Up Program did not change parental beliefs because of targeting. [18]

Finally, we must emphasize that the implementation of the Jamaica Home-Visiting Program, the 3T Home Visiting Intervention, and the LENA Start Program occurred in research settings that are human-capital abundant. In contrast, the Reach Up implementation in Colombia occurred within a real-world setting and explored human resources available to local

[18] See also Sylvia et al. (2021).

communities. For example, the home visitors were mothers who resided in the same community. However, we remark that the Reach Up implementation in Bangladesh (Hamadani et al., 2006) and China (Heckman et al., 2020) also used resources available in the community. While the latter two studies did not measure parental beliefs, they produced impacts on language and cognition much greater than Colombia's and aligned with the Jamaica Home Visiting Program's impacts. Carneiro et al. (2019) evaluated the NEP-I program implemented with community resources and found positive impacts on beliefs. Nonetheless, these authors document that the Chilean government spent resources recruiting and training the NEP-I workforce.

These contrasting findings suggest another explanation. The Reach Up Program did not change parental beliefs because it was implemented with low fidelity. Unfortunately, we cannot rule out this explanation because no data measures the quality of these programs when they occur in real-world settings. However, such data is crucial to advance the science of the determinants of early investments and the formulation of public policy that fosters human capital formation.

We conclude by suggesting two avenues of future research. First, this paper has established that the LENA Start Program impacted the language environment in the short run. Future research should investigate if this impact persists and, if so, whether it translates into improved language comprehension and literacy skills in the long run. Such research would inform the debate about the long-term impacts of early interventions (e.g., Bailey et al., 2020; García & Heckman, 2022).

Second, our LENA Start Program's implementation did not occur in a real-world setting. Thus, future research should investigate the LENA Start Program's implementation fidelity in a typical setting (e.g., as implemented by a school district or public library). If the execution is high quality, researchers should measure the impacts on parental beliefs and language outcomes in the short and long run. This research is vital to shedding light on vital scientific questions about the malleability of parental beliefs and the formulation of public policy to foster early human capital formation.

TABLE 1
Balance Test Results of the LENA Start Program

	Panel A: Baseline sample					
	Invitation to LENA Start		Control		p-value	
	(1)		(2)		(1) vs (2)	
VARIABLES	N	Mean	N	Mean	Robust	Clustered
Household income is less than or equal to twice the FPL[1]	142	0.704 (0.458)	145	0.752 (0.434)	0.368	0.345
Mother has attended some higher education program	144	0.097 (0.297)	145	0.103 (0.306)	0.861	0.906
Mother is Hispanic	144	0.153 (0.361)	145	0.200 (0.401)	0.294	0.354
Mother is Non-Hispanic black	144	0.625 (0.486)	145	0.579 (0.495)	0.429	0.477
Mother is Non-Hispanic white	144	0.139 (0.347)	145	0.159 (0.367)	0.639	0.689
Mother is single	144	0.729 (0.446)	145	0.655 (0.477)	0.174	0.263
Mother is cohabiting	144	0.063 (0.243)	145	0.083 (0.276)	0.509	0.454
Mother is married	144	0.208 (0.408)	145	0.262 (0.441)	0.283	0.373
Standardized HOME Score	133	-0.204 (1.383)	132	-0.044 (0.887)	0.265	0.203
Conversational Turns at 9 months	47	283.279 (191.053)	54	317.517 (120.901)	0.292	0.095
Adult Word Counts at 9 months	47	13761.77 (6981.438)	54	16260.48 (6804.080)	0.072	0.037
Seconds of Exposure to TV	47	7308.39 (5007.889)	54	7052.62 (4813.342)	0.795	0.727
Standardized Language Score from BSID[2]	137	-0.443 (0.708)	134	-0.502 (0.816)	0.521	0.469
Child currently being cared for by someone else regularly	131	0.718 (0.452)	132	0.659 (0.476)	0.308	0.398
Number of different regular childcare arrangements	131	1.092 (0.932)	132	0.985 (0.957)	0.360	0.401

TABLE 1
Balance Test Results of the LENA Start Program

	Panel B: Post-attrition sample					
	Invitation to LENA Start		Control		p-value	
	(3)		(4)		(3) vs (4)	
VARIABLES	N	Mean	N	Mean	Robust	Clustered
Household income is less than or equal to twice the FPL[1]	63	0.825 (0.383)	71	0.775 (0.421)	0.466	0.428
Mother has attended some higher education program	65	0.031 (0.174)	71	0.070 (0.258)	0.291	0.404
Mother is Hispanic	65	0.138 0.348	71	0.183 (0.390)	0.482	0.394
Mother is Non-Hispanic black	65	0.631 (0.486)	71	0.606 (0.492)	0.765	0.813
Mother is Non-Hispanic white	65	0.123 (0.331)	71	0.155 (0.364)	0.594	0.542
Mother is single	65	0.831 (0.378)	71	0.676 (0.471)	0.036	0.026
Mother is cohabiting	65	0.062 (0.242)	71	0.070 (0.258)	0.836	0.828
Mother is married	65	0.108 (0.312)	71	0.254 (0.438)	0.026	0.032
Standardized HOME Score	61	-0.351 (1.429)	68	-0.015 (0.686)	0.097	0.067
Conversational Turns at 9 months	25	278.407 (214.376)	36	327.929 (115.985)	0.295	0.210
Adult Word Counts at 9 months	25	14958.98 (7966.382)	36	17536.44 (6942.454)	0.195	0.132
Seconds of Exposure to TV	25	8272.94 (5481.453)	36	6798.43 (4631.373)	0.275	0.204
Standardized Language Score from BSID[2]	64	-0.484 (0.667)	71	-0.505 (0.789)	0.872	0.881
Child currently being cared for by someone else regularly	60	0.700 (0.462)	67	0.672 (0.473)	0.733	0.762
Number of different regular childcare arrangements	60	1.100 (1.053)	67	1.015 (0.961)	0.637	0.726

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

[1]Federal Poverty Line.

[2]Bayley Scale of Infant Development.

TABLE 2
Impact of the LENA Start Program on Conversational Turns

	Dependent Variable: Conversational Turns		
Variables	Model 1	Model 2	Model 3
	Panel A: OLS Estimator		
Random assignment to control or invitation arm	0.130	0.215	0.314**
Clustered SE	(0.187)	(0.160)	(0.113)
Robust SE	(0.160)	(0.163)	(0.152)
Randomization inference p-value	0.462	0.257	0.072
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***
Clustered SE	(0.286)	(0.211)	(0.125)
Robust SE	(0.268)	(0.264)	(0.233)
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

NOTE. — Standard errors in parentheses. We report clustered and robust standard errors, and we report the most conservative significance level. Clustered standard errors are clustered at the recruiment group level. We also report the randomization inference p-value after 2,000 stratified clustered resampling. Panel A reports the estimates obtained from ordinary least squares (OLS) regressions. Panel B reports the estimates obtained from two-stage least squares regression (2SLS) regressions using treatment assignment indicator as an instrument for attendance. We add the following variables to control for differences in recording sessions. First, we control for recording duration in endline, a dummy variable for a recording done on Saturday, and another dummy variable for a recording done on Sunday (*Model 1)* . Second, we additionally add demographic characteristics (*Model 2)* including a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording; and dummy for randomization group (block). Third, we additionally add conversational turn counts at baseline, recoding duration at baseline, and dummies for Saturday and Sunday at baseline (*Model 3).*

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

TABLE 3

Impact of the LENA Start Program on 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
Clustered SE	(0.155)	(0.178)	(0.177)
Robust SE	(0.162)	(0.153)	(0.169)
Randomization inference p-value	0.012	0.479	0.174
Observations	87	87	87
R-squared	0.749	0.692	0.679
	2SLS Estimator		
Attendance to the LENA Start program	0.690***	0.203	0.432*
Clustered SE	(0.172)	(0.228)	(0.234)
Robust SE	(0.230)	(0.215)	(0.245)
Observations	87	87	87
R-squared	0.745	0.706	0.675

TABLE 3
Impact of the LENA Start Program on Initiation of Conversations in Audio Segments with Focus Child and an Adult

	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
Clustered SE	(0.191)	(0.217)	(0.202)
Robust SE	(0.169)	(0.167)	(0.181)
Randomization inference p-value	0.002	0.214	0.111
Observations	87	87	87
R-squared	0.728	0.563	0.624
	2SLS Estimator		
Attendance to the LENA Start program	0.862***	0.414	0.572**
Clustered SE	(0.209)	(0.256)	(0.223)
Robust SE	(0.245)	(0.238)	(0.256)
Observations	87	87	87
R-squared	0.726	0.592	0.637

NOTE. — Standard errors in parentheses. We report clustered and robust standard errors, and we report the most conservative significance level. Clustered standard errors are clustered at the recruiment group level. We also report the randomization inference p-value after 2,000 stratified clustered resampling. Panel A reports the effects on the number of audio segments. Panel B reports the effects on the number of conversational turns by audio segment initiator. For both outcome measures, we report estimates obtained from ordinary least squares (OLS) regressions and two-stage least squares regression (2SLS) regressions using the treatment assignment indicator as an instrument for attendance. In all the regressions, we add the following variables to control for differences in recording sessions. We add the lagged dependent variable and we control for recording duration in baseline and endline. 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).

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

TABLE 4
Investigation of Potential Mechanisms of the LENA Start Program

VARIABLES	Parental Self-Efficacy	Parental Sense of Social Support	Parental Knowledge	Parental Beliefs	
				Error-Ridden	Factor Score
	Panel A: OLS Estimator				
Random assignment to control or invitation arm	-0.030	-0.077	0.293	0.306*	0.451**
Clustered SE	(0.158)	(0.153)	(0.185)	(0.151)	(0.176)
Robust SE	(0.131)	(0.176)	(0.139)	(0.154)	(0.143)
Randomization inference p-value	0.834	0.665	0.037	0.069	0.003
Observations	128	128	128	128	128
R-squared	0.569	0.352	0.614	0.318	0.377
	Panel B: 2SLS Estimator				
Attendance to the LENA Start program	-0.057	-0.146	0.564**	0.586**	0.859***
Clustered SE	(0.258)	(0.240)	(0.264)	(0.246)	(0.215)
Robust SE	(0.226)	(0.304)	(0.238)	(0.276)	(0.242)
Observations	128	128	128	128	128
R-squared	0.569	0.357	0.645	0.312	0.410

NOTE. — Standard errors in parentheses. We report clustered and robust standard errors, and we report the most conservative significance level. Clustered standard errors are clustered at the recruiment group level. We also report the randomization inference p-value after 2,000 stratified clustered resampling. Panel A reports the estimates obtained from ordinary least squares (OLS) regressions. Panel B reports the estimates obtained from two-stage least squares regression (2SLS) regressions using treatment assignment indicator as an instrument for attendance. In all the regressions, we add the following variables to control for differences in survey responses. We add the lagged dependent variable. Additionally, we add a dummy for maternal year of birth between 1978 and 1987; dummy for Hispanic ethnicity; dummy for non-Hispanic black; dummy for single mother; dummy for cohabiting mother; dummy for income below two times the federal poverty line; dummy for mothers with some postsecondary education; dummy for male child; age of the child at the date of the recording; and dummy for randomization group (block).

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

TABLE 5
Mediation Analysis Results

	Dependent Variable: Conversational Turn Counts		
		Mediators	
	ITT	ITT + Parental Beliefs	
		Error-Ridden	Factor Score
	(1)	(2)	(3)
Random assignment to control or invitation arm	0.314**	0.253*	0.163
Clustered SE	(0.113)	(0.137)	(0.147)
Robust SE	(0.152)	(0.148)	(0.145)
Maternal beliefs (error-riden)		0.098	
Clustered SE		(0.094)	
Robust SE		(0.085)	
Maternal beliefs (factor score)			0.198**
Clustered SE			(0.074)
Robust SE			(0.099)
Observations	90	90	90
p-value Joint Significance of mediators (clustered)			
p-value Joint Significance of mediators (robust)			
Control Group Mean	-0.089	-0.089	-0.089
Monte Carlo 95% CIs for Indirect Effects (clustered)		[-0.043; 0.168]	[0.026; 0.288]
Monte Carlo 95% CIs for Indirect Effects (robust)		[-0.035; 0.169]	[0.002; 0.321]

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

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

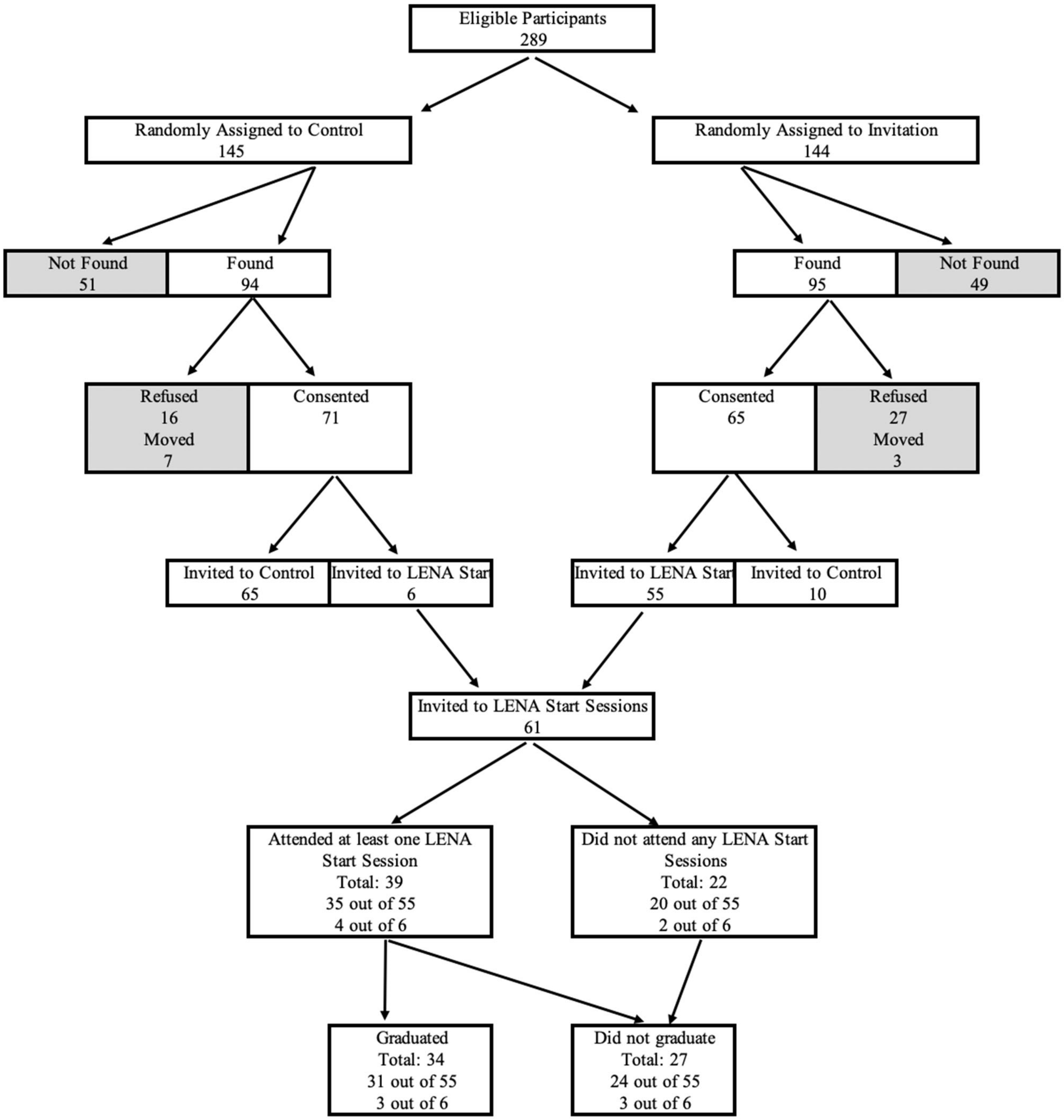

Figure 1. — LEME Study: Eligibility, Random Assignment, Recruitment, Consent, Final Assignment, Graduation.

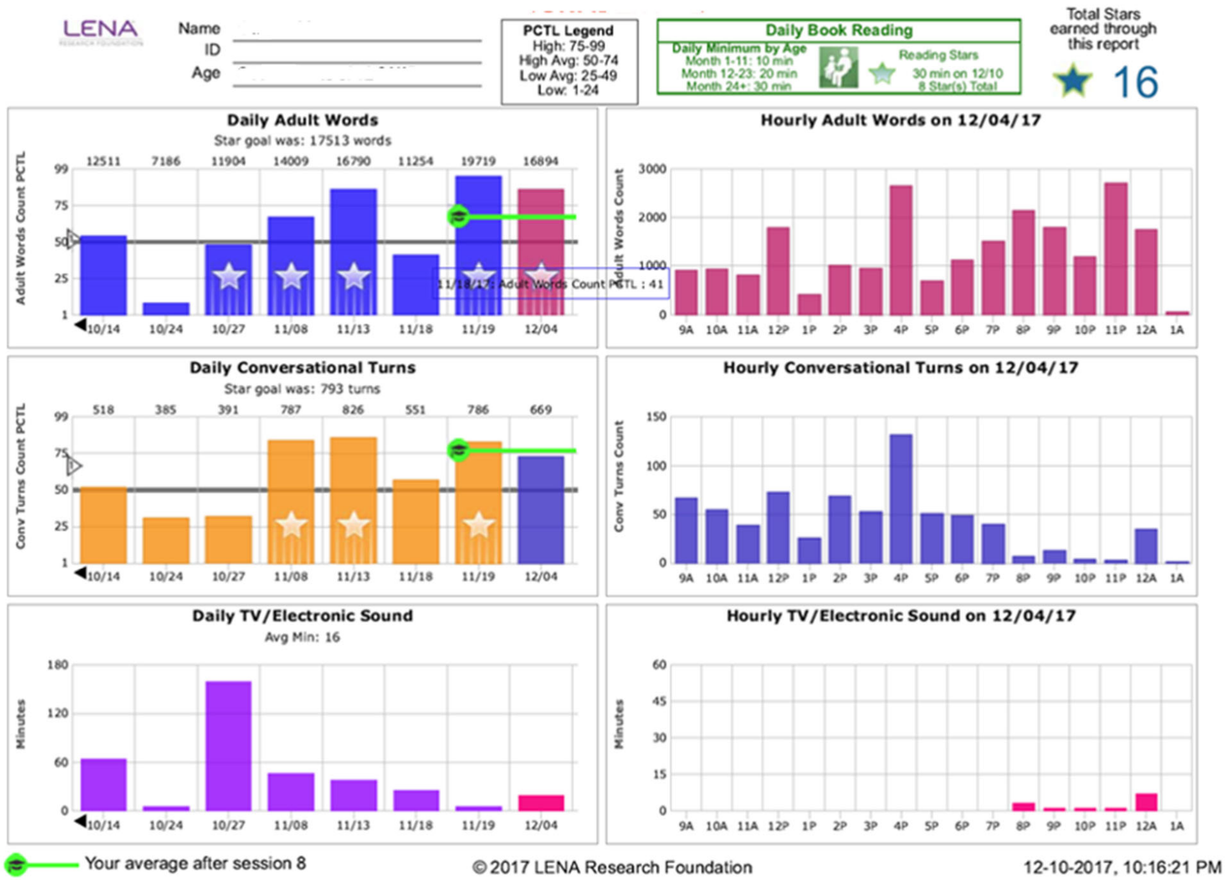

Figure 2. — LENA Start Objective Feedback Report.

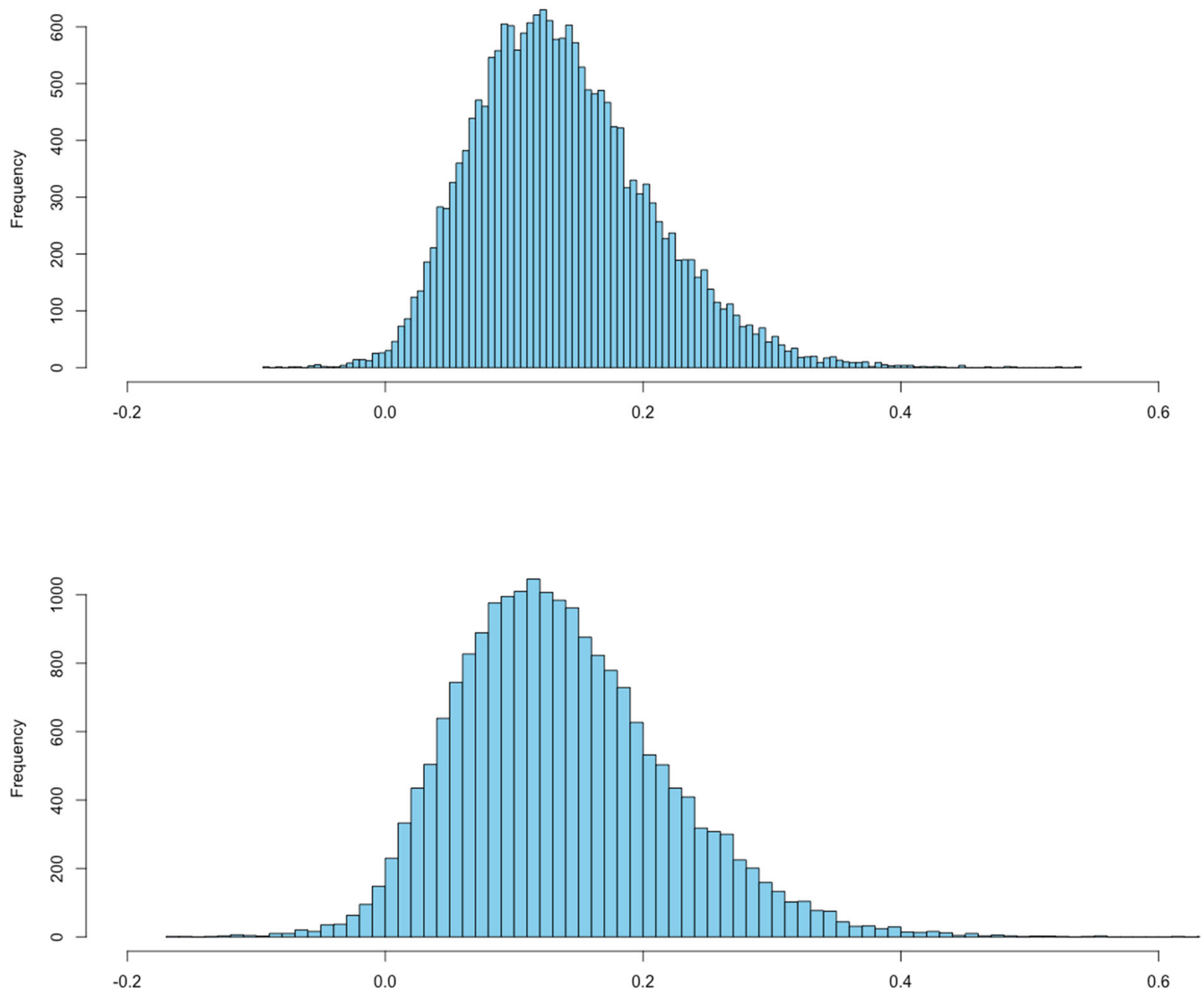

Figure 3. — Distribution of the Indirect Effect of the factor score of parental beliefs, clustered (top) and robust (bottom) standard errors.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Suskind, Dana L. , Kristin R. Leffel, Eileen Graf, Marc W. Hernandez, Elizabeth A. Gunderson, Shannon G. Sapolich, Elizabeth Suskind, Lindsey Leininger, Susan Goldin-Meadow, and Susan C. Levine. "A Parent-Directed Language Intervention for Children of Low Socioeconomic Status: A Randomized Controlled Pilot Study." *Journal of Child Language* 43, no. 2 (2016): 366-406.

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

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

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

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

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

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

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

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

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

Appendix

A. Related Literature on Children's Language Development

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

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

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

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

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

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

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

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

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

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

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

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

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

B. The Context and The Philadelphia Human Development Study

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

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

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

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

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

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

The third visit happened at the Children's Hospital of Philadelphia research building when the child was between twenty-two and twenty-six months old. We assessed the development of 674 (out of the 822) children using the Bayley Scale of Infant Development (BSID III). The BSID III contains two language development dimensions: receptive language (ability to understand language input) and expressive language (ability to produce language output). First, we average these two scores to estimate a composite language score. Then, we use the BSID III normalizing statistics to calculate the age-normed score. Table B1 shows that the children in the inner-city sample are 56% of a standard deviation below the mean in language development. In comparison, the children in the suburban sample are 21% of a standard deviation above the mean. Therefore, we estimate a difference of nearly 77% of a standard deviation in language development around age two years.

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

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

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

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

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

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

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

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

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

In summary, many (but not all) children growing up in low-income households have deficient language development, and the deficits in language development correlate with gaps in the language environment measured one year earlier. Furthermore, in the PHD Study data, the language environment's heterogeneity – measured by conversational turn counts – has a fragile association with family income. For these reasons, our study hypothesizes that parental beliefs about the importance of the early language environment for a child's language development drive parental linguistic input. Through the LENA Program, we investigate if it is possible to change parental linguistic input and, if so, whether parental beliefs are one of the mechanisms of this change.

C. Results of the Recruitment for the LENA Start Program

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

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

Third, we discuss the attendance results at the LENA Start Program. Similar to previous exercises, let $d_i^A = 1$ if our team successfully made an eligible study participant to attend at least one session, conditional on consenting to participate in the study, and $d_i^A = 0$, otherwise. Again, we estimate a linear probability model in which d_i^A is the dependent variable, and the random assignment is the only independent variable. Table C1 shows that the random assignment to control or invitation arms was the most crucial variable in predicting whether a family attended at least one of the LENA Start Program sessions. Families randomly assigned to the LENA Start Program invitation arm were likelier to attend the sessions. This result also holds when we do not condition on having consented. It indicates that the random assignment is not a weak Instrumental Variable.

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

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

D. Results on Construction of the Recording Data

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

E. Additional Details on Survey Measures

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

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

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

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

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

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

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

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

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

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

Table E3 examines the balance between control and treatment arms at baseline. Table E4 shows that the impact in the aggregate scale masks heterogeneity in the subscales. When we disaggregate the scales, the point estimate of the ITT suggests relatively larger impacts on “Reading Books,” “Learning Math,” and the “School Readiness” subscales and more negligible impacts on “Learning to Talk” or “Language and TV” subscales. The point estimate of the LATE parameter suggests impacts on “Learning Math” and the “School Readiness” subscales but no impacts on the other three scales.

Parental Beliefs: We ask parents the following question:

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

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

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

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

F. Robustness Analysis

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

G. Additional Results on Adult Words and Exposure to TV

Now, we discuss the impact of the LENA Start Program on adult words. The LENA System produces several counts of adult words. First, it automatically generates an overall adult word count. This automatic count combines words addressed to the child and speech overheard by the child. Because the literature emphasizes the role of speech addressed to the child in events of joint attention or speech recasting, we use the ADEX software to isolate adult words in audio segments that have conversations between the focus child and a female or male adult. By imposing this additional constraint, we increase the likelihood that the adult words are part of child-directed speech. As we have done so far, we use the lagged dependent variable models described in equations (1) and (2)-(3), but the results are the same if we use the FE or IV-FE procedures described in equations (9) and (10)-(11).

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

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

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

Next, we also investigated the impact of the LENA Start Program on the focus child's exposure to audio from the TV or other electronics. Table G2 shows the results. As in other tables in the paper, Panel A reports the ITT estimates, while Panel B documents the LATE estimates. We do not find any impact of the program on exposure to TV. The point estimates in Model 1 indicate an increase. The sign of the estimates turns negative (denoting a reduction in exposure to TV) once we control for demographic characteristics (Model 2). If we also include statistics about TV/Electronics at baseline, the coefficient becomes even smaller, but it is still statistically insignificant. We conclude that the LENA Start Program does not reduce exposure to TV/Electronics.

Table B1: Characteristics of PHD Study by Residence

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table C2: First-stage Results of the LATE Strategy

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

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

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

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

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

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

Table D2: Conversational Turn Counts at Baseline

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

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

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

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

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

Table E2: Results of Item Response Theory Analysis

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

	Fixed Effect Specification			
	(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*
Clustered SE	(0.072)	(0.063)	(0.063)	(0.075)
Robust SE	(0.179)	(0.176)	(0.177)	(0.174)
Observations	214	214	214	214
Number of participants	122	122	122	122
	Panel B: 2SLS Estimator			
Attendance to the LENA Start program	0.578*	0.582**	0.600**	0.580*
Clustered SE	(0.152)	(0.142)	(0.149)	(0.167)
Robust SE	(0.301)	(0.297)	(0.298)	(0.299)
Observations	214	214	214	214
Number of participants	122	122	122	122

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

Table G1

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

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

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

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

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

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

TABLE G3
Mediation Analysis Results

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

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

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Suskind, Dana L. , Kristin R. Leffel, Eileen Graf, Marc W. Hernandez, Elizabeth A. Gunderson, Shannon G. Sapolich, Elizabeth Suskind, Lindsey Leininger, Susan Goldin-Meadow, and Susan C. Levine. "A Parent-Directed Language Intervention for Children of Low Socioeconomic Status: A Randomized Controlled Pilot Study." *Journal of Child Language* 43, no. 2 (2016): 366-406.

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

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

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

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

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

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

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

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

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