Deeper than Demographics: Using Predictors of High-quality Language Environments Can Maximize Limited Resources

Key Findings from Alief ISD Early Language Development Study

and Its Implications for Initiative Design

April 13, 2020

RICE | SOCIAL SCIENCES

Executive Summary

In the classic children's picture book *Brown Bear, Brown Bear*, a narrator asks a repetitive, open-ended question of a variety of animals: "What do you see?" Each animal responds by describing another animal, and the book follows a predictable rhythm. The exchange, however simple, illustrates one of the most critical building blocks of early childhood development: a conversational turn.

When we consider the greatest needs in early childhood research and the design and delivery of programs, "serve and return" parent-child interactions are a valuable measure of what is happening inside a home. Demographics and socioeconomic labels such as "low-income" are often considered predictors of early childhood development, but these categories are too broad, and fail to capture the specific parental behaviors that help children cultivate vibrant vocabularies and high developmental capital. Ultimately, what a parent does is more important than who the parent is.

Responsive relationships between parent and child are essential to a child's brain development, but Harris County lacks data on the quality of language environments. Thus, we have begun to assemble a fact base to measure early language development, offering a more detailed view of behaviors, predictors, and the striking differences between families that share socioeconomic and demographic characteristics. We begin with nearly 500 3-year-old old children of low-income families in southwest Harris County in the Alief Independent School District from 2016-2018. We collected objective measures of "serve and return" parent-child interactions, recorded by the LENA device, a small wearable device — often referred to as a "talk pedometer." Among the findings:

- About 40% of children are in households with low-quality language environments (see figure 1). 46% of the children growing up in a household with a low-quality language environment are developmentally delayed.

- About 20% of low-income families provide high-quality levels of language environment. About 73% of the children who are growing up in low-income families and who excel in measures of "serve and return" are on track to be school-ready.

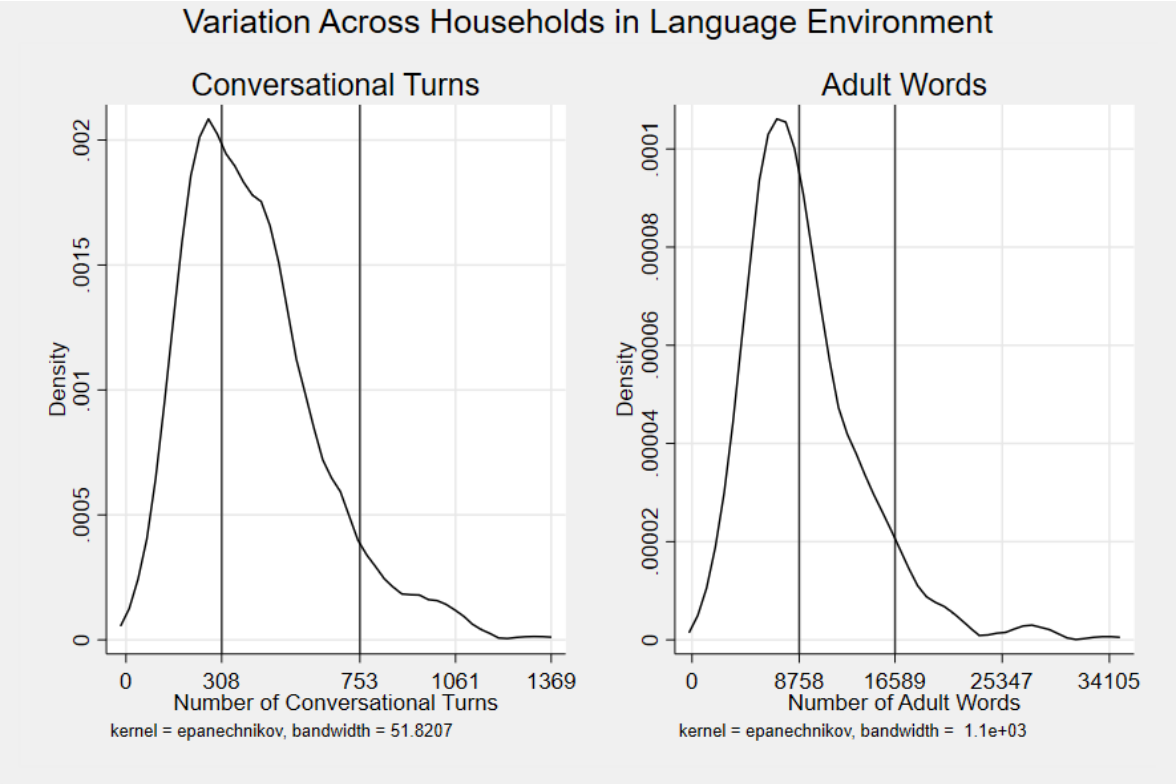
- Striking differences between low-income families in the same community are driven by what parents do, not who they are. Some low-income children in Alief will have heard 6.6 million fewer words by age five than other low-income children in the same community.

If, in the case of these Alief families, we aim to target children experiencing the lowest levels of conversational turns, then the target population is about 40% of children in low-income families. Thus, it is important to screen families on measures that predict parental behaviors, such as:

- Time devoted to meaningful speech
- Child's exposure to TV/electronics

While most early childhood interventions do not screen families in this way, it is feasible to test such an approach. Doing so would deploy limited resources to toddlers who are least likely to enter school ready to learn. A significant fraction of low-income children are failing to reach even basic levels of literacy because they do not have appropriate levels of language comprehension skills. By investing in this group of children, whose future developmental capital hinges on the quality of their language environment, we can improve educational outcomes while generating the greatest returns to society.

Figure 1



The area between the two black vertical lines contain 50% of the observations in the nationally representative data. The children in these households experience average levels of language environment as their conversational turns are between 308 and 753 and adult words are between 8,578 and 16,589 adult words per 12 hours.

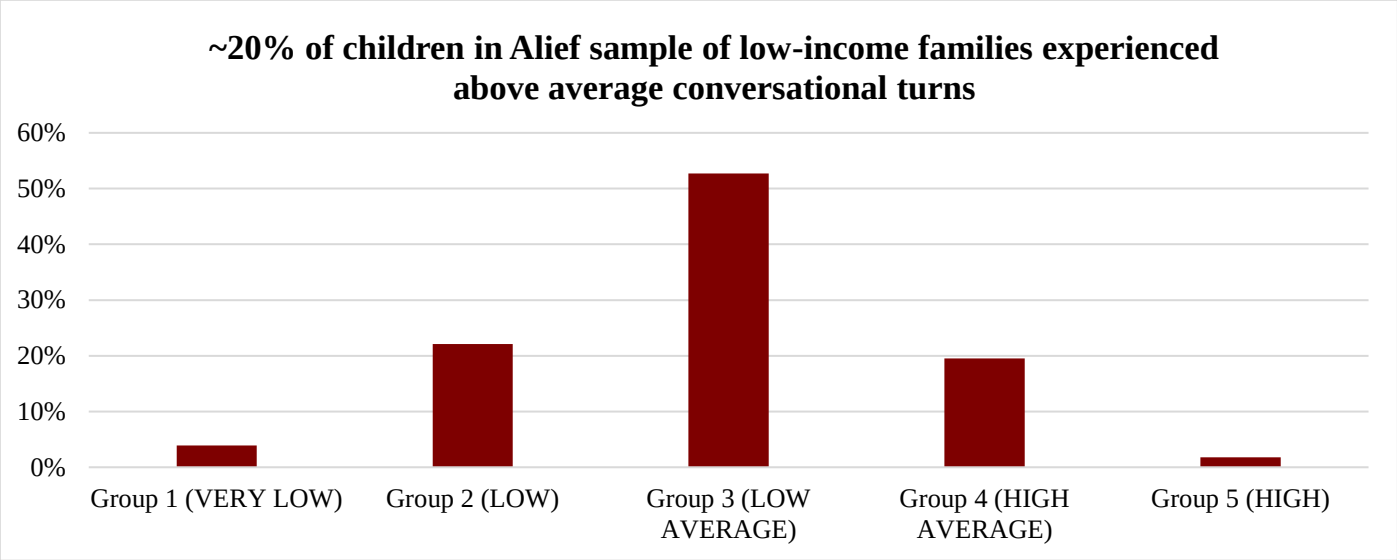
Overall, we see a low-quality language environment among Alief families. In our dataset, nearly 40% of the households have fewer than 308 conversational turns and 55% have fewer than 8,758 adult words. At the same time, ~20% of these households have above average conversational turns. To illustrate the variation across low-income families, we organize families into five groups based on the total number of conversational turns in a 12 hour period (holding constant meaningful speech and TV time):

- Group 1 (VERY LOW): Up to 171 conversational turns per 12 hours.
- Group 2 (LOW): 172 to 308 conversational turns per 12 hours.
- Group 3 (LOW AVERAGE): 309 to 506 conversational turns per 12 hours.

- Group 4 (HIGH AVERAGE): 507 to 703 conversational turns per 12 hours.
- Group 5 (HIGH): 704 or more conversational turns per 12 hours.

In our Alief sample, we find the following distribution of children across each of the five groups based on level of conversational turns above:

Figure 2



A. Striking Differences Between Low-income Families Driven by Parental Behavior

A body of sound research makes clear the importance of reading at grade level by third grade for improving educational and labor market outcomes. Literature also suggests that early language environment in the first three years of a child's life lays the groundwork for reading and other outcomes, with one study indicating that conversational turns (child speaks and an adult responds, or vice versa) between adults and young children accounted for 14% to 27% of the variance in later dimensions of human capital[1].

Previous studies have shown socioeconomic differences in the language environment, with a significant fraction of low-income children failing to reach even basic levels of literacy because they do not have appropriate levels of language comprehension skills.

Low-income households are a diverse group, and language environments are not dictated by demographic characteristics. One low-income household may display high numbers of conversational turns, just as a high-income household may display a lower quality language environment. By focusing on parental associated with a high number of conversational turns, early childhood education initiatives can provide targeted support only to those who need it – maximizing limited resources.

The heterogeneity observed in the language environment of low-income families is driven more so by what parents do, and less so on who parents are (demographic characteristics). For example, parental education was the demographic variable most predictive of conversational turns. Even so, adjusting parental behaviors such as increasing meaningful speech and decreasing exposure to TV/electronics predicts an increase in conversational turns that is almost six times the predictive change due to parental education.

These large differences driven by parental behavior imply that designing and delivering programs targeting low-income families as a homogenous group runs the risk of serving families with high-quality language environments while missing those most in-need of support. Not all

[1] Human capital is the combination of all attributes – intrinsic to an individual – that has or produces economic value. It includes all forms of skills (e.g., literacy and numeracy skills) as well as knowledge (whether formal or tacit), creativity, health conditions, and personality attributes (e.g., patience, risk aversion/loving, grit). What makes human capital special is that it is intrinsic to an individual and is not transferrable

children growing up in low-income households are at-risk of experiencing low-quality language environments. Additionally, if the within-group inequality we document for language environment is also true for other aspects of parenting – such as the establishment of secure attachment or fostering the formation of executive functioning skills – then parenting programs need to spend more resources in screening to make sure that these programs are reaching at-risk families within low-income segments of our society.

B. Language Development, Reading Skills Are Force Multipliers in Children's Future Ability to Grow and Prosper

The results from the 2017 National Assessment of Educational Progress (NAEP, 2017) report a sizable socioeconomic gap in reading skills. In the United States, 46% of low-income[2] children enrolled in fourth grade have reading skills below the basic level. In contrast, the same rate for children coming from middle- and high-income households is about 18%.

Research shows that the capacity to read proficiently confers positive economic returns in the labor market. For example, Hanushek, Schwerdt, Wiederhold, and Woessmann (2015) use the Programme for the International Assessment of Adult Competencies (PIAAC) data to estimate the return to numeracy, literacy, and problem-solving skills. They find that one standard deviation improvement in literacy skills raises hourly wage rates by approximately 7% even after controlling for potential experience, educational attainment, problem-solving skills, and numeracy skills. Reading proficiency is so fundamental in modern life that it is considered a basic human right (UNESCO, 1975).

Research also suggests that early language environment influences early language comprehension skills and language development more broadly. Weisleder and Fernald (2013) show that the quality of a child's language environment[3] at baseline (age 19 months) predicts

[2] By low-income children we mean children whose family income is low enough that they are eligible for the National School Lunch Program.

[3] 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.

vocabulary and language processing skills[4] five months later, even after controlling for vocabulary and language processing skills at baseline. Gilkerson et al. (2018) report the results of a longitudinal study that first measures the quality of the language environment of infants and toddlers[5] and then when the same children are 9 to 14 years old, measures their IQ and language development.[6] The authors show that conversational turns between adults and children accounted for 14% to 27% of the variance in later dimensions of human capital.

C. There Are Large Differences in Language Environment by Socioeconomic Status

Research shows that there is a sizable socioeconomic gradient in the language environment. The relation between family socioeconomic status (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, while children from low-SES families were exposed to about 616 per hour. Rowe and Goldin-Meadow (2009) found that high-SES mothers gesture more about objects in the environment when they are close to their infant and toddlers. They also report that lower SES mothers not only talk less often, use smaller vocabulary, and employ syntactic structures that are less varied or complex, they also tend to talk to direct their children's behavior, not so much to converse and engage with their children. Higher SES children are more likely to be exposed to rarer vocabulary, more linear narratives, more open-ended questions, and other characteristics of speech more closely aligned to the academic language environment of the school system.[7]

All in all, it is a desirable goal to have children become proficient readers, but a significant fraction of low-income children are failing to reach even basic levels of literacy because they do

[4] To measure language processing skills, the authors use the Looking-While-Listening task (LWL, Fernald et al., 2008). In the LWL procedure, infants look at pairs of pictures while listening to speech naming one of the pictures, and their gaze patterns are video-recorded as the sentence unfolds in time. Language processing skills is then measured as reaction time, that is, the amount of time the infant shifts away from the distracter to the target picture. Children with higher processing skills take less time to shift away to the target picture.

[5] 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.

[6] 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).

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

not have appropriate levels of language comprehension skills. The development of these skills requires greater exposure to language so that children can practice and hone their language processing skills and simultaneously increase their vocabulary.

Gathering objective measures of child-parent interaction in Harris County is critical for understanding how to improve early childhood outcomes.

In this document, we analyze a unique dataset that we collected in a partnership with the Alief Independent School District in Houston, TX. Specifically, we collected objective measures of "serve and return" with nearly 500 low-income families in academic years 2016-2017 and 2017-2018. To measure the quality of the child-parent interaction, our team assessed "serve and return" with 489 children in the Alief ISD zone with the LENA System. Families who agree to participate place the LENA recording device in appropriate clothing, which the child wears for a day. The availability of these data provides insights about the quality of the language environment of vulnerable children in one specific geographic area of Harris County. The LENA device allows us to measure "serve and return" interactions through five key measures:

- **Conversational Turn Counts:** Among these measures, conversational turns provide the best indicator of back and forth interactions between children and caregivers. It provides an estimate of the total number of conversational turns the child engages in with an adult per amount of time. The LENA System uses similar rules to Hart and Risley (1995) to measure conversational turns. A conversational turn is considered to have occurred when a child vocalizes and an adult responds, or an adult speaks and a child responds.
- **Adult Word Counts:** Adult word count (AW) is an estimate of the number of adult words spoken near a child per a given amount of time.
- **Child Vocalization Counts (CV):** Child vocalization is language expression by children. It differentiates from AW because young children vocalize but do not necessary say recognizable words. Importantly, the LENA-based algorithms can distinguish between key child speech and key child non-speech. Sounds that were considered speech included words, babbles, and pre-speech communicative sounds or "protophones" such as squeals, growls, or raspberries.

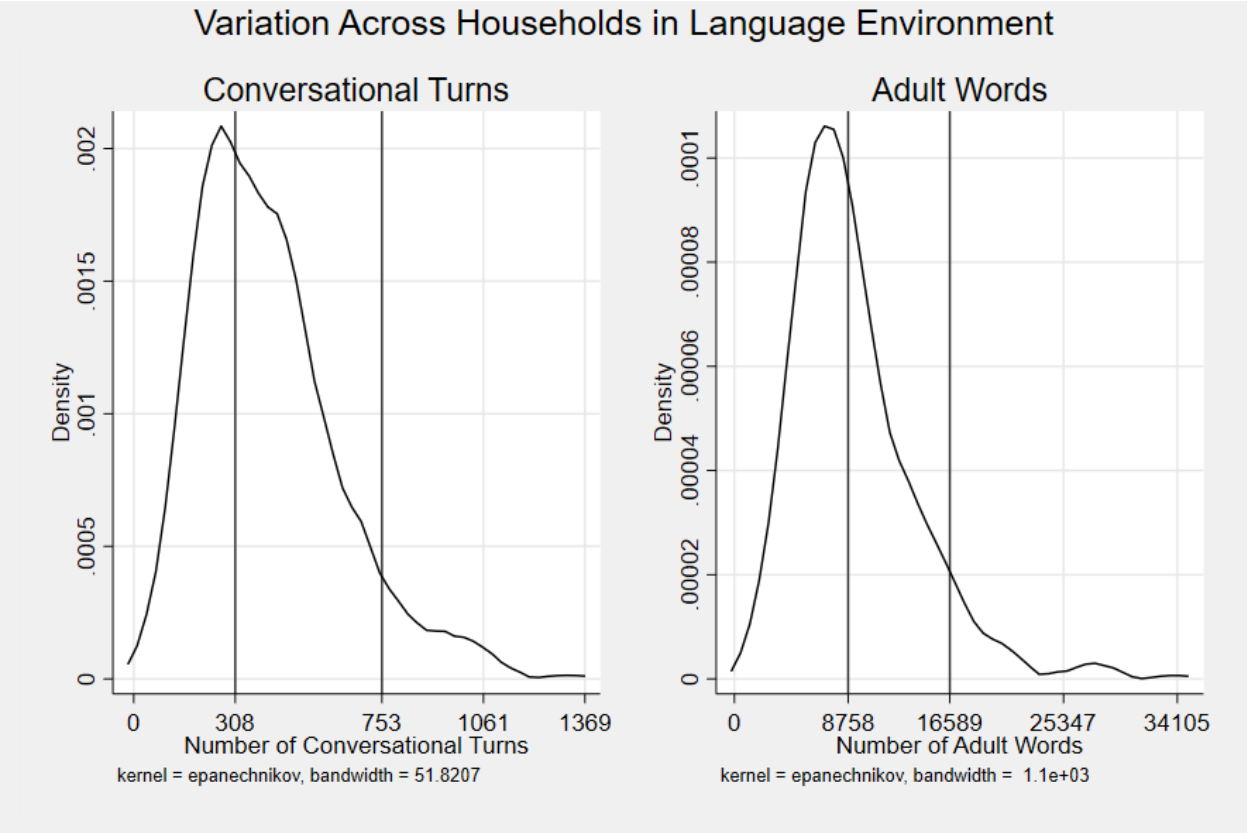
- **Meaningful Speech Time (MS):** Usable, distinguishable speech that is included in the reported information. The fact that speech is distinguishable implies that the adult person is close to the child. It does not imply that the adult is directly interacting with the child because the adult could be watching the child play while talking to a friend on the cell phone. It does, however, provide an estimate of the amount of time that the adult person is near the key child.
- **TV/Electronics Exposure Time (TV):** Audio specifically from a television or other electronic sources.

'Serve and return' results for families in Alief sample indicate ~40% of low-income households have below average levels of conversational turns.

We focus on language and audio environment statistics aggregated to a period of 12 hours. Figure 1 displays the density of conversational turns (left) and adult words (right) within our sample. Each one of the density curves are marked with black vertical lines that indicate the 25th and 75th percentiles according to the national norms as estimated by the LENA Foundation. In the nationally representative data, the area to the left of the first black vertical line contains 25% of households. The language environment in these households is low because the children experience at most 308 conversational turns or hear at most 8,758 adult words during a period of twelve hours.

In contrast, the area to the right of the second black vertical line contain 25% of the observations in the nationally representative data. The households to the right of the second black line have high levels of language environment because they have at least 753 conversational turns or hear 16,589 adult words during 12 hours.

Figure 1



Finally, the area between the two black vertical lines contain 50% of the observations in the nationally representative data. The children in these households experience average levels of language environment as their conversational turns are between 308 and 753 and adult words are between 8,578 and 16,589 adult words per 12 hours.

Overall, we see a low-quality language environment among Alief families. In our dataset, nearly 40% of the households have fewer than 308 conversational turns and 55% have fewer than 8,758 adult words. To clarify what this means, suppose we had a representative sample of 100 families from Texas and assume that low-income families are similar to low-income families in Alief. Because this sample is representative of the State, 34 families in our sample would be low-income.[8] Out of the 100 representative families, 25 families would have fewer than 308 conversational turns. However, 14 of these families would be low-income families and 11 would

[8] Family income below 200% of the Federal Poverty Line.

be from families with income above the poverty threshold. Therefore, even though low-income families represent only 34% of a representative sample in Texas, they represent 56% of the families with low levels of conversational turns.

Lower levels of conversational turns predict developmental delay. While the sample includes only low-income families in Alief, we see large disparities in developmental delay.

Next, we investigate the extent to which conversational turns predict measures of language development, which we assess in two different ways. First, we explore the information on language development as produced by the LENA system. In particular, LENA estimates CV (the number of words, babbles, and "protophones" or prespeech communicative sounds produced by the child).

Additionally, we used the Bracken School Readiness Assessment – Third Edition (BSRA-3) to assess the children's preschool readiness. BSRA-3 includes five subtests to assess basic concepts related to school readiness, and all of the subtests reflect receptive and expressive language skills. The subjects include Colors, Letters, Numbers/Counting, Size/Comparison, and Shapes. For the purposes of our analysis, we classify children as "delayed" – if their BSRA-3 score is at or below the 17th percentile for the age. This cutoff is the standard for the BSRA-3 instrument.

To analyze the relationship between conversational turns and developmental delays as measured by BSRA-3, we organize families into five groups based on the total number of conversational turns in a 12 hour period (holding constant meaningful speech and TV time):

- Group 1 (VERY LOW): Up to 171 conversational turns per 12 hours.
- Group 2 (LOW): 172 to 308 conversational turns per 12 hours.
- Group 3 (LOW AVERAGE): 309 to 506 conversational turns per 12 hours.
- Group 4 (HIGH AVERAGE): 507 to 703 conversational turns per 12 hours.
- Group 5 (HIGH): 704 or more conversational turns per 12 hours.

In our Alief sample, we find the following distribution of children across each of the five groups based on level of conversational turns above:

Figure 2

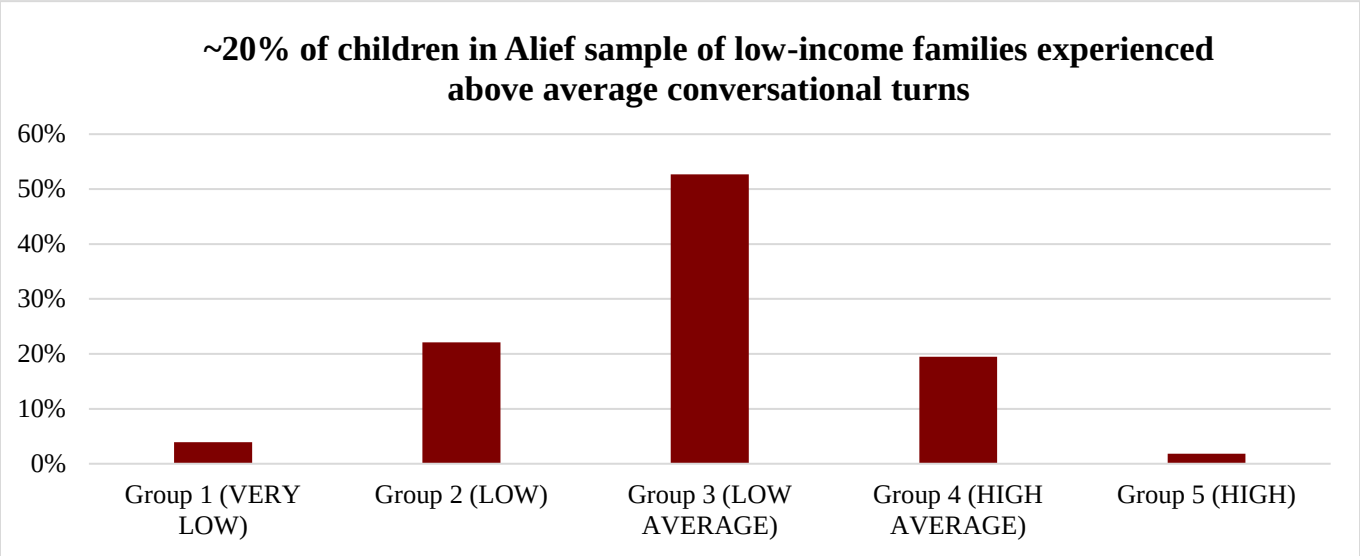
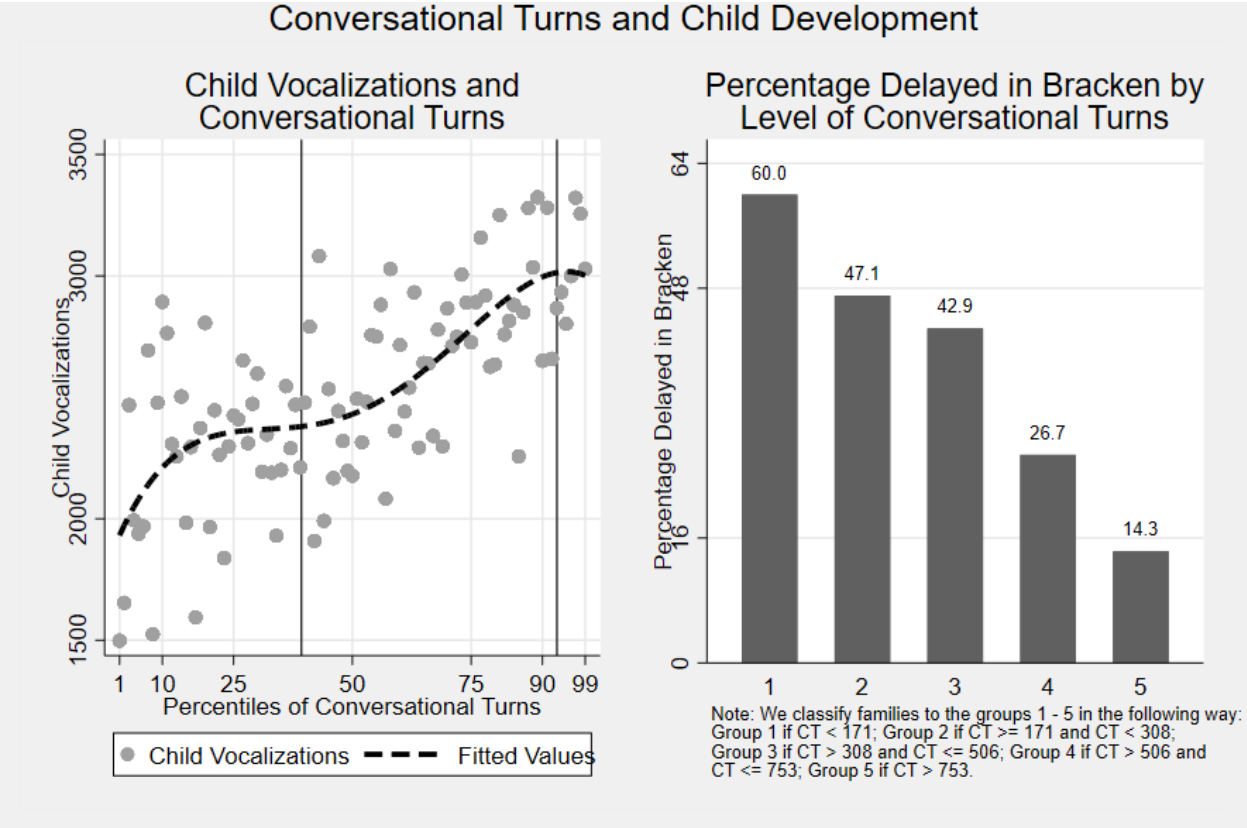


Figure 3 plots the correlation between conversational turns and child vocalizations (left) and the relationship between conversational turns and developmental delay (as measured by BSRA-3, right). To produce these graphs, we control for meaningful speech and TV time for conversational turns and child vocalizations. By controlling for meaningful speech and TV, we are sure that the correlation only captures the direct relationship between conversational turns and child vocalizations, and not the indirect relationship through meaningful speech or TV time. For BSRA-3, we control for the age at the time of testing. Therefore, the relationships depicted in Figure 3 capture variation in conversational turns holding constant other determinants of child vocalizations and scores in BSRA-3.

Figure 3



There is a positive relationship between conversational turns and child vocalizations. Children with low levels of conversational turns (with percentiles to the left of the first vertical line) have approximately 2,000 vocalizations a day, while the children with high levels of conversational turns express approximately 3,000 vocalizations a day (or 50% more). This finding is concerning, as research shows that children with fewer than 2,000 child vocalizations per twelve hours of the day have much higher risks of experiencing language developmental delays as measured by reliable instruments of language development (Gilkerson and Richards, 2009).

The results denote a clear and strong negative correlation between conversational turns and developmental delay as indicated by the BSRA-3. In Group 1, the group with the lowest number of conversational turns, 60% of the children are classified as delayed (their score in BSRA-3 is below the 17th percentile for age in the nationally representative data). This number is lower, but still very high, for Groups 2 and 3. About 27% of the children in Group 4 are delayed, which represents 45% of the share of children delayed in Group 5. Finally, 14% of the children in Group

5 are developmentally delayed. The huge disparity in developmental delay across conversational turn groups – within a narrowly defined socioeconomic population – indicates a need to understand parental behaviors that predict outcomes for children because differences in school readiness are not driven by basic demographic characteristics (e.g. income level).

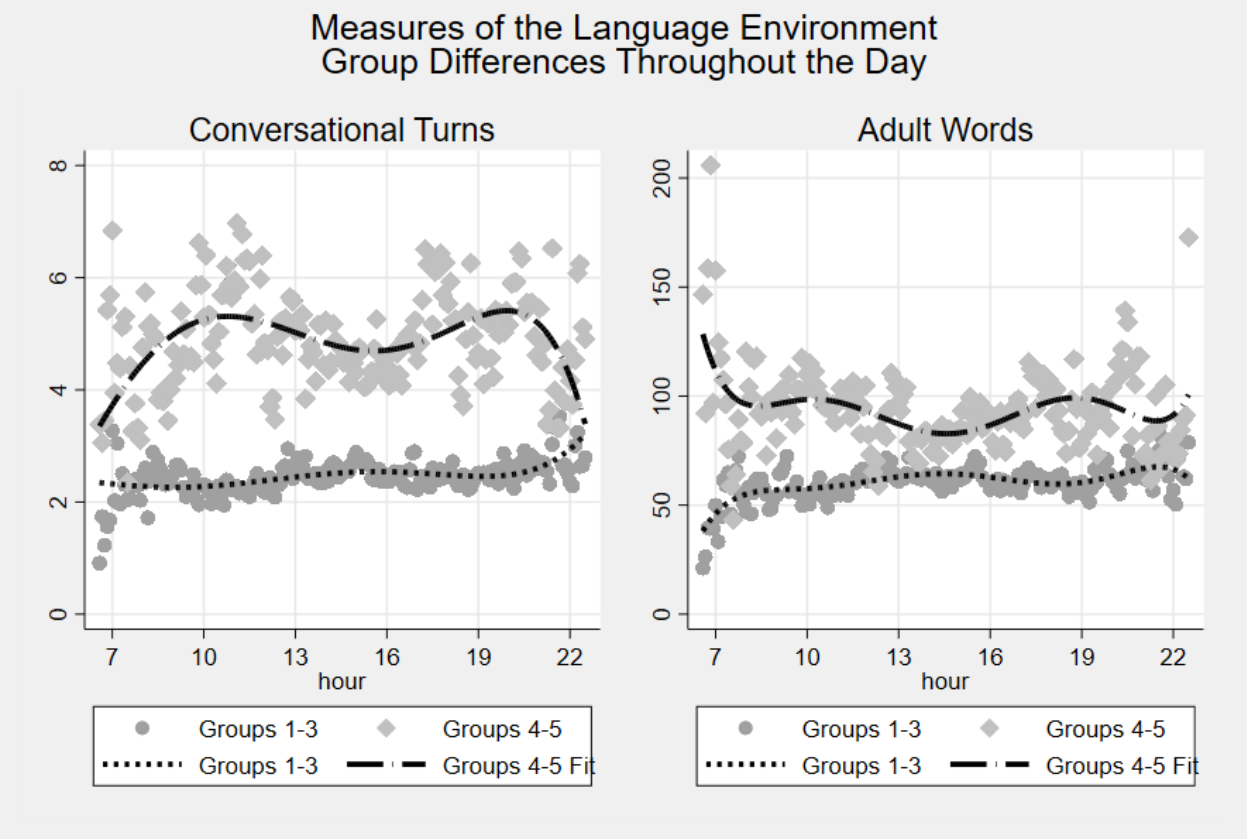
Low-income households are a diverse group when it comes to language environment. Some low-income children in Alief will have heard 6.6 million fewer words by age 5 than other low-income children in the same community.

All in all, the literature to date has focused on differences across socioeconomic groups (e.g., Hartley and Risley, 1995). The data from Alief ISD show that there is large heterogeneity within low-income families. Our findings demonstrate that not all low-income families are identical in terms of the quality of the language environment. For example, about 7% and 8% of the sample are in high levels of conversational turns and adult words, respectively. Furthermore, 53% and 37% of the Alief families are in the average range of language environment. This finding implies that not all children growing up in low-income households are at risk of experiencing low-quality language environments. If, for example, we aim to target children in the low group of conversational turns, then the target population is about 40% of the children in low-income families.

Figure 4 further emphasizes the important diversity of language environments within low-income families. Figure 4 compares conversational turns and adult words between Groups 1-3 and 4-5 throughout the day. Because we use detrended data, the differences between groups are not driven by natural variation common to all groups, but rather variation that is specific to each group. The differences in Language Environment are clear: children in Groups 4-5 tend to have 2.5 more conversations on average per five-minute segment. Although the difference seems small, there are 144 such segments during the period of 12 hours. Thus, a small number at a five-minute segment transforms into a difference of nearly 366 conversational turns at the end of the period of 12 hours. If the across-group differences we report are estimates that reflect language environment on typical days for each child, then at the end of the child's first four years of life, children in Groups 1-3 will have a little over half a million fewer conversational turns with adults than children in Groups 4-5. We document the existence of these huge gaps in conversational turns within a homogeneous population in vulnerability as measured by socio-demographic variables.

The right panel of Figure 4 compares AW across the two groups throughout the day. Although the number of AW fluctuates throughout the day for Groups 4-5, the differences between groups are more or less constant during the entire day. Within a five-minute segment, children in Groups 4 and 5 hear about 31 more adult words than the children in Groups 1-3. Once we aggregate this difference to a period of 12 hours, the difference is above 4,500 adult words. Again, assuming this is a typical day for the child, the daily deficits in adult words translate into a four-year gap of over 6.6 million words.

Figure 4

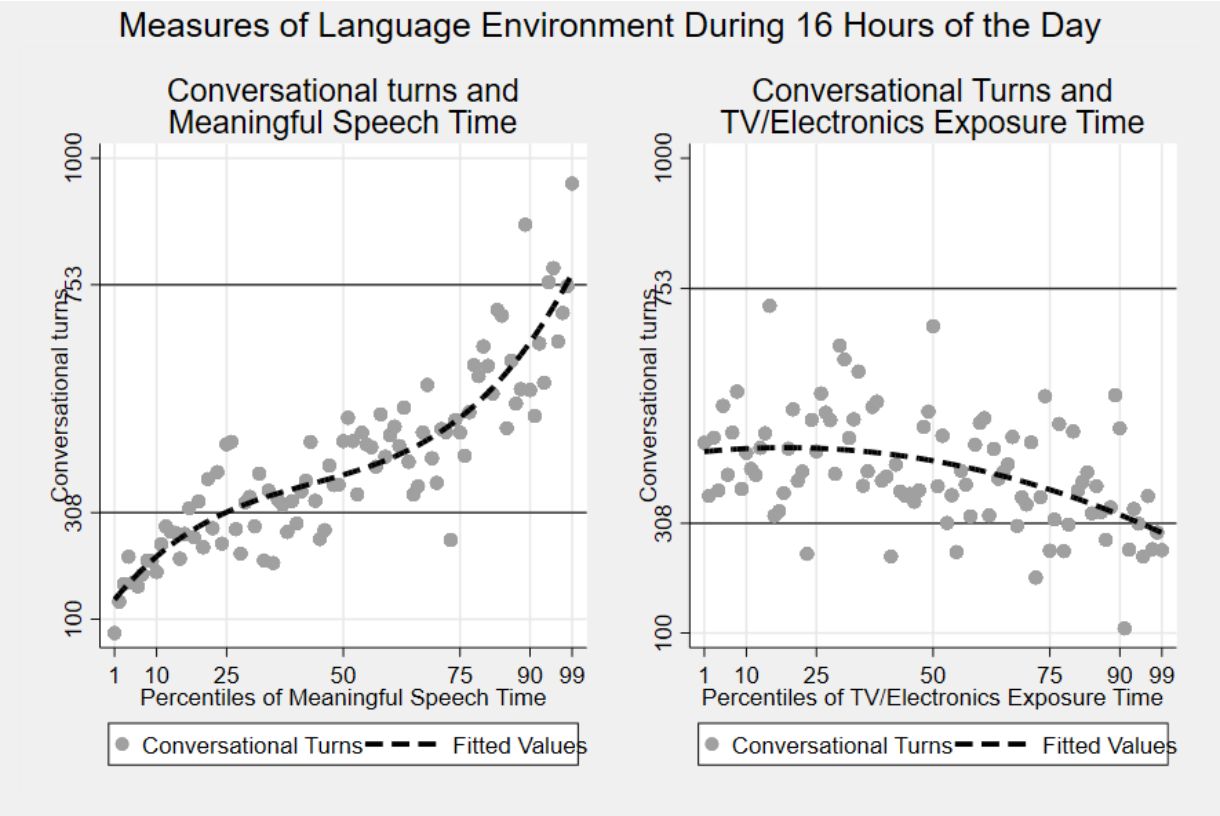


D. Heterogeneity In the Language Environment Driven by What Parents Do, Rather Than Who They Are

We analyze the determinants of the variability in language environment across households in this disadvantaged sample. In particular, we investigate if the amount of time that the parent spends with the child (as measured by Meaningful Speech Time - MS) or the amount of time that the child is exposed to TV or other electronics (as measured by TV/Electronics Time - TV) explains any of the variation in the quality of the language environment. The covariation between MS and CT is not necessarily mechanic. For example, a parent may be very close to the child (and not necessarily paying attention to the child) while talking to a friend on the telephone or another adult person in the room. This action increases MS, but it would does not lead to higher CT. Thus, we focus on CT instead of AW to make sure that we are capturing events of adult-child interactions.

In Figure 5, the horizontal axis displays the percentiles of MS (left) and TV (right). The vertical axis displays the counts of conversational turns. To help with interpretation, we added two horizontal lines that separate the groups of "low," "average," and "high" levels of CT. Figure 5 shows that, the higher the amount MS, the higher the number of CT. Families with low amounts of MS are families at-risk for falling into low levels of CT and, in contrast, families at the top of MS are families more likely to reach high levels of CT.

Figure 5



In contrast, the relationship between TV and CT is negative. The higher the exposure to TV or electronics, the higher the chance that the CT fall in the "low" region level. Therefore, if improvements in language development require increasing CT between adults and children, it is necessary to increase MS and to decrease exposure to TV or other electronics.

Designing and delivering programs targeting low-income families as a homogenous group runs the risk of serving families with high-quality language environments, while missing those most in-need of support. It is important to understand factors determining heterogeneity in language environment within low-income families so that we target and identify the at-risk children that may benefit most from an intervention.

This insight on targeting becomes even more important because our sample is homogenous in terms of socio-economic vulnerability. This homogeneity makes it impossible to find demographic variables that can explain differences in CT in usual ways as the vast majority of the

households in our sample are poor and constituted by parents from minority racial or ethnic groups. There is, however, one dimension of heterogeneity that is predictive of CT: parental education.

Table 1 compares the associations between conversational turns and a dummy variable that is equal to one if the parent has some post-secondary education (i.e., highest grade completed is above 12 years) or not. Holding constant MS and TV, parents with a post-secondary education have 40 more conversational turns than parents that do not have a post-secondary education. This association is 33% larger than the association between "low" and "high" TV exposure, where "low" indicates that the child is at the 25th percentile of exposure and "high" indicates that the child is at the 75nd percentile. However, the association between CT and MS is about four times the association between CT and parental education (164 vs 40). A simultaneous increase in MS and decrease in TV predicts an increase in conversation turns that is almost six times the predictive change due to parental education.

Table 1

Associations Between Conversational Turns, Parental Education, Meaningful Speech Time and TV/Electronics Exposure Time[1]

	Parent does not have post-secondary degree[2]	Parent has post-secondary degree[2]	Difference	p-value
Conversational Turns	422	462	40	0.047
	TV/Electronics Time is Above 25th Percentile	TV/Electronics Time is At 25th Percentile	Difference	p-value
Conversational Turns	411	439	29	0.000
	Parent's Meaningful Speech Time is Below 75th percentile	Parent's Meaningful Speech Time is at 75th Percentile	Difference	p-value
Conversational Turns	371	535	164	0.000
	Meaningful Speech Time is Below 75th Percentile *and* TV/Electronic Time is Above 25th Percentile	Meaningful Speech Time is At 75th Percentile *and* TV/Electronic Time is At 25th Percentile	Difference	p-value
Conversational Turns	360	591	231	0.000

Notes: [1]To produce the figures in this table, we regress conversational turns on percentiles of meaningful sppech time, percentiles of TV/Electronics exposure time, a dummy variable that is equal to one if the parent has post-secondary education and zero otherwise, child's age at the time of measurement and the a dummy for boys. To produce the estimates of conversational turns, we vary the variable of interest and set the meaningful speech time and/or TV/Electronics exposure time to their mean values. The estimates of conversational turns for variations in Meaningul Speech Time and/or TV/Electronics time assume that the parent does not have a post-secondary education.

[2]The parent has a post-secondary education if the parent reports that her highest grade completed is above 12 years of education.

E. Screening and Targeting Based on Key Predictors of Parental Behavior Is Critical for Serving Low-income Families with Greatest Needs

The above finding emphasizing the importance of parental behaviors compared to parental demographic characteristics suggests that the quality of parent-child interaction, at least as captured by the quality of language interaction, is strongly associated with what parents do and less so to who the parents are. Again, this implies that targeting should not focus on demographic or socioeconomic variables, but on measures that predict parental behaviors given demographic and socioeconomic characteristics.

The finding on the importance of increasing meaningful speech and decreasing TV/electronic exposure is important for identifying at-risk children, as it indicates the importance of screening families on measures that predict parental behaviors (e.g., within low-income families, screen: parent use of time, time spent with children, child exposure to TV). Time-use surveys do not directly assess MS, but they measure the amount of time parents spend with the child (whether interaction was the primary activity or not). However, it is unknown how parental reports of time with the child match MS as measured by the LENA System. Thus, a pilot that simultaneously carried out an assessment of the language environment and asked parents to report their use of time (or of their children's) could (a) link parental reports to MS as measured by LENA and (b) use this information to develop a screening instrument focused on parental behaviors that could improve targeting. Additionally, parenting surveys ask parents to report the amount of time children are exposed to TV or other electronics.

Therefore, the combination of parental reports in these two dimensions could produce a low-cost screening instrument that identifies families that could benefit the most from interventions aimed to improve the quality of the children's language environment. In spite of its importance, no such study has been carried out. While most early childhood interventions do not screen families in this way, it is feasible to test such an approach. Doing so would deploy limited resources to toddlers who are least likely to enter school ready to learn.