This is the *Accepted Manuscript* of an article published by the American Speech-Language-Hearing Association (ASHA) in the *Journal of Speech, Language, and Hearing Research*. The manuscript is reprinted here with permission from ASHA and is further available online https://doi.org/10.1044/2023_JSLHR-23-00069

The Stability and Validity of Automated Indices of Vocal Development in Infants with Autistic and Non-Autistic Siblings

Jennifer E. Markfeld, MS[a]* & Jacob I. Feldman, PhD[b,c]*,

Claire Daly, MS[a,d], Pooja Santapuram, MD[e,f], Sarah M. Bowman, MD[b,g], Kacie Dunham-Carr, BA[a,h,i], Evan Suzman, MS[j,k], Bahar Keçeli-Kaysılı, PhD[b], Tiffany G. Woynaroski, PhD[b,c,h,l]

[a] Department of Hearing & Speech Sciences, Vanderbilt University, Nashville, TN, USA

[b] Department of Hearing & Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

[c] Frist Center for Autism and Innovation, Vanderbilt University, Nashville, TN, USA

[d] Present Affiliation: Children's Healthcare of Atlanta, Egleston Hospital, Atlanta, GA, USA

[e] Vanderbilt University School of Medicine, Nashville, TN, USA

[f] Present Affiliation: Department of Anesthesiology, Columbia University Irving Medical Center, New York City, NY, USA

[g] Present Affiliation: Department of Pediatrics, Cincinnati Children's Hospital, Cincinnati, OH, USA

[h] Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA

[i] Neuroscience Graduate Program, Vanderbilt University, Nashville, TN, USA

[j] Master's Program in Biomedical Science, Vanderbilt University, Nashville, TN, USA

[k] Present Affilitation: UT Southwestern School of Medicine, Dallas, TX

[1] Vanderbilt Kennedy Center, Vanderbilt University Medical Center, Nashville, TN, USA

* Authors made equal contributions to this work.

Author Note

Jennifer E. Markfeld, MS, CF-SLP, is a PhD student in the Department of Hearing and Speech Sciences at Vanderbilt University. Jacob I. Feldman, PhD, CCC-SLP, is a research fellow in the Department of Hearing and Speech Sciences at Vanderbilt University Medical Center. Claire Daly, MS, was a master's of speech-language pathology student in the Department of Hearing and Speech Sciences at Vanderbilt University Medical Center at the time this work was completed and is presently a Clinical Fellow in Speech-Language Pathology at Children's Healthcare of Atlanta. Pooja Santapuram, MD, was a medical student in the School of Medicine at Vanderbilt University at the time this work was completed and is presently a resident in anesthesiology at Columbia University Irving Medical Center. Sarah M. Bowman, MD, was previously a research assistant in the Department of Hearing and Speech Sciences at Vanderbilt University Medical Center and is presently a resident in pediatrics at Cincinnati Children's Hospital. Kacie Dunham-Carr, BA, is a PhD student in Neuroscience and Hearing and Speech Sciences at Vanderbilt University. Evan Suzman, MS, was a graduate student in the Master's Program in Biomedical Science at Vanderbilt University at the time this work was completed and is presently a medical student at UT Southwestern School of Medicine. Bahar Keçeli-Kaysili, PhD, is a Clinical and Translational Research Coordinator in the Department of Hearing and Speech Sciences at Vanderbilt University Medical Center. Tiffany Woynaroski, PhD, is Assistant Professor in the Department of Hearing and Speech Sciences, Investigator in the Vanderbilt Kennedy Center at Vanderbilt University Medical Center, Fellow of the Frist Center for Autism and Innovation at Vanderbilt University, and member of the Vanderbilt Brain Institute at Vanderbilt University.

Correspondence regarding this manuscript may be addressed to Jennifer Markfeld via email at jennifer.e.markfeld@vanderbilt.edu, by phone at (804) 551-1510, or by mail at MCE 8310 South Tower, 1215 21st Avenue South, Nashville, TN 37232.

Abstract

Purpose: This study evaluates the extent to which automated indices of vocal development are stable and valid for predicting language in infants at increased familial likelihood for autism and/or language impairment and relatively lower likelihood infants.

Method: A group of infants with autistic siblings (20 infants; Sibs-autism) and a comparison group of infants with non-autistic siblings (20 infants; Sibs-NA) wore Language ENvironment Analysis (LENA) recording devices for 16 hours on two days within a one-week period. Extant software was used to derive several putative indices of vocal development from these recordings. Stability of these variables was examined across and within groups. Expressive and receptive language aggregates were calculated for each participant. Multiple regression analyses were used to (a) evaluate zero-order correlations for variables derived from LENA recordings with concurrent and future language and (b) test whether those associations were moderated by group status.

Results: Both stability and validity differed by variable and group status. All variables reached acceptable stability in the Sibs-autism group within two to three observations, while stability of most variables was attenuated in the Sibs-NA group. No variables were associated with concurrent language in the theoretically-motivated direction across groups, but two variables were strongly associated with concurrent expressive language in only the Sibs-NA group. Additionally, two variables were associated with later expressive language, though these correlations were again stronger in the Sibs-NA versus Sibs-autism group.

Conclusions: Although selected automated indices of vocal development were stable in Sibs-autism and/or valid for predicting expressive language within Sibs-NA, no scores showed strong, theoretically-motivated associations with language within the Sibs-autism group. Automated indices of vocal development may, thus, have limited validity or clinical utility for predicting language development in infants at elevated familial likelihood for autism.

Keywords: vocal development, LENA, autism

The Stability and Validity of Automated Indices of Vocal Development in Infants with Autistic and Non-Autistic Siblings

Prelinguistic vocalizations are recognized as the foundation upon which language is built. Infants, universally, pass through distinct vocal stages enroute toward spoken word and broader spoken language use (Oller, 2000), and aspects of prelinguistic vocal development such as the frequency, duration, complexity, and reciprocity of vocalizations appear to be useful for predicting language development in a number of clinical and at-risk populations (Jensen et al., 1988; Patten et al., 2014; Plumb & Wetherby, 2013). Measuring early vocalizations may be especially useful for young children with or at heightened likelihood for autism, who have been reported to experience early disruptions in vocal development and to have highly variable language outcomes (e.g., Chericoni et al., 2016; Paul et al., 2011; Sheinkopf et al., 2000; Swanson et al., 2018; Tager-Flusberg et al., 2005; Tager-Flusberg & Kasari, 2013; Yankowitz et al., 2022).

A large body of literature has found that autistic preschool children present with altered or reduced preverbal vocalizations (e.g., Sheinkopf et al., 2000; Wetherby et al., 1988; Wetherby et al., 1998) that may predict later expressive language (McDaniel et al., 2019; Woynaroski et al., 2016; see McDaniel et al., 2018 for a review). A few studies have specifically shown that vocal development in infancy and toddlerhood may predict future autism features and language skills in young autistic children. For example, the occurrence of syllabic, speech-like vocal productions in the second year of life were previously observed to predict pervasiveness of future autism features (Plumb & Wetherby, 2013), and indices of vocal development obtained from standardized assessments in toddlerhood were found to predict later expressive language outcomes (Chawarska et al., 2007). However, most of this research has been carried out using conventional behavior sampling approaches, which are extremely time-intensive (and thus costly) to collect and code. As a result, such measures likely have limited potential to translate to use in everyday clinical practice.

Due to recent technological advances, a wide range of variables purported to index vocal development can now be derived from automated analysis of audio recordings collected via Language ENvironment Analysis (LENA) recorders (Gilkerson et al., 2017). The LENA device is a small recorder that fits into specialized clothing worn by the child. LENA devices have a 16-hour recording capacity, allowing for long-form audio recordings of a child's home language environment. The accompanying software allows for the derivation of many indices purported to tap child vocal development.

Indices currently available for commercial use through LENA standard packages include: Child Vocalization Count (CVC), Child Vocalization Duration (CVD), and Automated Vocal Analysis (AVA) raw scores (based on the distribution of biphone pairs; see Richards et al., 2017 for further detail). Additional proprietary variables, such as the Infraphonological Vocal Development (IVD) and Average Count Per Utterance (ACPU) scores, can be obtained through the LENA Foundation for research purposes (Xu et al., 2014), and still others such as the Reciprocal Vocal Contingency score (Harbison et al., 2018) can be derived via external software programs. Although these indices provide valuable information about a child's vocal development, LENA cannot provide precise information on aspects of language such as mean length of utterance or linguistic diversity for children or adult speech (Putnam et al., 2023). Additionally, there is limited evidence to support its utility for school-aged autistic children with limited spoken language (e.g., Jones et al., 2019; Woynaroski et al., 2017).

Several studies have evaluated the psychometrics of variables that can be derived via automated vocal analysis in preschool-aged autistic children (e.g., Bredin-Oja et al., 2018; Harbison et al., 2018; Warren et al., 2010; Woynaroski et al., 2017; Yoder et al., 2013). For example, automated indices of vocalization count and duration were reported to differentiate groups of preschoolers with and without autism and to be sensitive to short-term treatment effects (Warren et al., 2010). Using generalizability (G) and decision (D) studies (Cronbach et al., 1963; Yoder et al., 2018), Woynaroski, Yoder, and colleagues found that several automated metrics of vocal development, including indices of vocal complexity and vocal reciprocity derived from LENA recordings, were highly stable (Harbison et al., 2018; Woynaroski et al., 2016). Further, some of these automated scores have been observed to predict concurrent and future expressive language in autistic preschoolers (e.g., Harbison et al., 2018; Trembath et al., 2019; Woynaroski et al., 2017; Yoder et al., 2013). Evidence for associations between automated indices of vocal development and language to date, however, is mixed (McDaniel et al., 2020; Rankine et al., 2017; Sulek et al., 2022; Woynaroski, 2014).

The work reviewed above provides some preliminary support for the consideration of automated vocal analysis in research, and perhaps ultimately in clinical practice, in autistic preschoolers. However, little is known at present about the application of automated indices of vocal development in infant siblings of autistic children, who are known to be at heightened likelihood for receiving a future diagnosis of autism and/or language impairment (Sibs-autism; Messinger et al., 2013; Ozonoff et al., 2014). Given that language development is highly heterogenous in Sibs-autism, and that early language intervention may be particularly important due to an increased likelihood for not only autism but also language impairment, measuring early vocal features of this population could inform future research and clinical practice (e.g., Hampton & Rodriguez, 2022; Ozonoff et al., 2014; Swanson et al., 2018). Some studies have used LENA to measure early vocal development in Sibs-autism; however, it is not yet known whether variables derived via this novel, automated approach are similarly stable and valid for predicting concurrent and/or future language in this population. Thus, it is not yet clear whether this novel technology may be useful for determining which infants within this high likelihood group might benefit from early intervention.

A recent systematic review of the use of LENA in autism research identified only a few published articles that used LENA in the Sibs-autism population (Putnam et al., 2023). Of these articles, one study examined the concurrent validity of LENA variables with other, standardized measures of language, and two others examined the predictive validity of LENA variables with later standardized measures of language (Markfeld et al., 2022; Seidl et al., 2018; Swanson et al., 2018). Further, these studies did not examine the full range of possible variables that can be derived from LENA when evaluating concurrent or predictive validity (e.g., a study evaluated CVC but no other child variables). Therefore, although concurrent and predictive validity between selected LENA variables purported to tap vocal development and language have been examined in Sibs-autism, the stability and validity of such scores is not yet well-established. However, the validity of a given measure is limited by its reliability, especially reliability related to the stability of estimates across contexts and observations (Crocker & Algina, 1986; Yoder et al., 2018). Thus, the stability of automated indices derived from LENA must be evaluated in Sibs-autism in order to determine whether these indices are potentially valid for predicting concurrent and/or future language. Additionally, the full range of child variables that can be derived from LENA recordings have not yet been assessed for concurrent and predictive validity with other language measures commonly used in Sibs-autism.

The present study sought to extend the aforementioned work to evaluate the extent to which several indices of vocal development that can presently be derived via automated vocal analysis are stable and associated with concurrent language in Sibs-autism, as well as the extent to which these indices predict language as assessed 9 months later. Specific research questions were as follows:

1. How many audio recordings are necessary to obtain stable automated indices of vocal development in Sibs-autism and infants at relatively lower, general population level likelihood for autism (i.e., Sibs-NA; infant siblings of non-autistic children)?
2. To what extent are automated indices of vocal development associated with concurrent language in Sibs-autism and Sibs-NA?
3. To what extent are these novel, automated indices valid for predicting future language in Sibs-autism and Sibs-NA?

Method

Recruitment and study procedures were carried out in accordance with the approval of the Vanderbilt Institutional Review Board. All caregivers provided written informed consent prior to their child's participation in the study.

Participants

Participants were 40 infants enrolled in a longitudinal study of language development, 20 Sibs-autism and 20 Sibs-NA (see Table 1; sample partially overlaps with previous work from our laboratory; i.e., Bottema-Beutel et al., 2019; Feldman et al., 2021; Santapuram et al., 2022). Participants were recruited via flyers, emails, local preschools and doctor's offices, word of mouth, and community outreach events at the local science center. All participants met the following inclusion criteria: (a) chronological age of 12-18 months (±30 days) at the time LENA recordings were collected and (b) living in a primarily English-speaking household. In the Sibs-autism group, at least one older sibling was diagnosed with autism in an evaluation that included a research-reliable administration of the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2012). For the Sibs-NA group, participants were required to (a) have only typically-developing siblings, as confirmed by screening below the threshold for autism risk on the Social Communication Questionnaire (Rutter et al., 2003) and (b) no first-degree relatives on the autism spectrum. Exclusion criteria for both groups were (a) adverse neurological history, (b) known genetic conditions, and (c) pre-term birth (gestation < 37 weeks). Groups did not differ on chronological age and biological sex but did differ on entry-level mental age (see Table 1). Additional participant demographics (i.e., race, ethnicity, and caregiver level of education as a proxy for socioeconomic status [SES]) are reported in Table 1.

Procedures

LENA data collection. All participants were provided with two LENA recording devices, which were worn for 16 hours each. This is the maximum recording time for LENA processors. Infants wore the LENA devices for two days within a one-week period in their natural environments (i.e., typical home and community settings). Additionally, participants were provided with a specialized garment (e.g., shirt, vest) to wear throughout recording sessions. Caregivers were instructed to turn the recorders on when their child woke up in the morning and to place the recorders in the front chest pocket of their child's garment. The devices ran continuously throughout the day, and the collected audio data was transferred onto password-protected laboratory computers upon return.

Across the two days of recordings, eight recordings were fewer than 16 hours long (*M* time of these eight recordings = 10.91 hours, *SD* = 1.96 hours, *Min-Max* = 9.34-14.97 hours).

Recordings were analyzed using the LENA Advanced Data EXtractor (ADEX) software to derive the following indices purported to tap child vocal complexity, frequency, and duration: Automated Vocalization Analysis (AVA) raw scores, Child Vocalization Count, and Child Vocalization Duration. Using modified speech algorithms, the LENA software segments audio data and categorizes each segment as likely to have been produced by the target child or by an alternate speaker/sound source. Segments identified as produced by the target child are further categorized as speech-related utterances, vegetative sounds, or fixed signals such as crying (for further detail regarding segmentation and classification, refer to Xu, Yapanel, & Gray, 2008; Xu, Yapanel, Gray, et al., 2008). AVA raw scores are based on bi-phone distributions within speech-related utterances, while Child Vocalization Count and Duration tap the frequency and duration of speech-related utterances, respectively (Richards et al., 2017; Warren et al., 2010).

A previously developed index of caregiver-child vocal reciprocity (Reciprocal Vocal Contingency score; RVC score; Harbison et al., 2018) was derived using extant software (Yoder et al., 2016). This index quantifies child vocal reciprocity based on bidirectional, three-event child-caregiver-child exchanges and controls for the chance sequencing of these vocal events.

Finally, we evaluated two additional proprietary LENA variables not presently included in the ADEX software suite. The Infraphonological Vocal Development (IVD) score quantifies vocal complexity based on 12 acoustic parameters purported to tap speech-likeness of child vocalizations. In brief, a software program presently available for research purposes is used to derive raw scores based on the presence or absence of the 12 parameters in "vocal islands" (syllable-like units) identified as being produced by the target child; raw scores are then weighted using beta weights from a regression model that predicted chronological age in a sample of typically developing children in Oller et al. (2010) to derive the final IVD scores (see Woynaroski et al., 2017; Yoder et al., 2013).

Average Count Per Utterance (ACPU) scores (Xu et al., 2014) were generated using open-source Sphinx recognition software that automatically identifies consonants, vowels, silence, and nonspeech sounds (such as lip smacking or coughs) within utterances labeled as being produced by the target child. We focused on the ACPU of the Sphinx-identified elements theoretically expected to predict future language - estimated phone counts. We derived both ACPU-Consonants and ACPU-Vowels scores. These scores were aggregated into an ACPU-Consonants+Vowels (ACPU-C+V) score because they are conceptually similar and were previously observed to be empirically related (Woynaroski et al., 2017).

Expressive and receptive language aggregates. To evaluate concurrent validity, infants' expressive and receptive language was evaluated at 12-18 months using the Mullen Scales of Early Learning (MSEL; Mullen, 1995), the Vineland Adaptive Behavior Scales (VABS; Sparrow et al., 2005), and the MacArthur-Bates Communicative Development Inventory: Words and Gestures (MCDI; Fenson et al., 2007). To evaluate predictive validity, the MSEL and VABS were also collected 9 months later (at Time 2), along with the Words and Sentences form of the MCDI.

At each timepoint, aggregates were generated for each participant by averaging the z-scores for (a) raw scores from the relevant indices of the MCDI and (b) the age equivalency scores from the relevant indices of the MSEL and VABS. Aggregates were used to enhance the stability and, thus, the potential construct validity of our language scores (Rushton et al., 1983), and have been used in prior work investigating language in autistic children as well as Sibs-autism (e.g., Feldman et al., 2021, Rogers et al., 2021).

Analytic Plan

All automated vocal indices were derived for the entire recording period. Prior to running analyses, Child Vocalization Count and Duration indices were divided by the total length of the recording to account for limited incomplete recordings (i.e., 8/80 recordings; 10% of sample). There is limited information from the LENA Foundation itself on recommended length of recording necessary to obtain reliable estimates of the indices derived in this study; however, available resources recommend at least one hour of recording for some indices (Gilkerson & Richards, 2020). Deriving rates of LENA indices is common and recommended by researchers who use LENA hardware and software as a method of reducing the potential for biased estimates in recordings that are likely to be influenced by recording duration (i.e., count variables) and may be incomplete (e.g., Bredin-Oja et al., 2018; Dykstra et al., 2012; Markfeld et al., 2023; Putnam et al., 2023). The remaining variables (i.e., AVA raw scores, RVC, ACPU-C+V, and IVD) are metrics unlikely to be influenced by the length of recording times due to how they are derived (i.e., by using proportions or means) and thus did not require correction, consistent with past studies of these automated measures (McDaniel et al., 2020; Seidl et al., 2018; Woynaroski et al., 2017; Xu et al., 2014; Yoder et al., 2013).

To answer our first research question, G and D studies (Cronbach et al., 1963; Yoder et al., 2018) were carried out for all automated indices of interest to evaluate the stability across and within groups. G studies quantified the test-retest reliability of variables based on how consistently children ranked relative to one another in terms of vocal development across repeated observations. These studies produced an intra-class correlation coefficient referred to as a *g* coefficient. Consistent with previous research (Bottema-Beutel et al., 2019; Sandbank & Yoder, 2014; Woynaroski et al., 2017), the a priori threshold for acceptable stability was set at *g*

≥ 0.8. D studies, which drew upon the variance estimates from the G studies to extrapolate beyond observed data, were used to determine the *g* coefficients up to a maximum of 6 days (i.e., 6 days with each day indexed by a maximum 16-hour-long LENA recording) and to assess which LENA variables reached our a priori stability threshold of $g \geq 0.8$ within this window.

To answer our second and third research questions, a series of multiple regression analyses was carried out to evaluate the magnitude of zero-order correlations between variables derived from audio recordings and (a) concurrent and (b) future expressive and receptive language, as well as to test whether the aforementioned associations were moderated by likelihood status. Any variables that were non-normally distributed (i.e., skew > |1| or kurtosis > |3|) were transformed, and missing data (ranging from 0-27.5% across variables) were imputed using the *missForest* package (Stekhoven & Bühlmann, 2012) in R (R Core Team, 2020). Interaction effects were probed as planned follow-up analyses at $p \leq .1$ due to our small sample size and the preliminary, exploratory nature of our planned analyses.

In a series of post-hoc analyses, we reran the above analyses while covarying for primary caregiver's level of education (see Hoff, et al., 2018; Huttenlocher et al, 2002; Justice, et al., 2020), which served as our proxy for SES, to evaluate whether results were robust to controlling for SES. We hypothesized that controlling for SES would not change our results.

A priori power analyses conducted with consideration of the magnitude of associations between LENA indices and language previously found for autistic children (e.g., Woynaroski et al., 2017) indicated that a minimum sample size of 40 was needed in order to detect zero-order correlations between LENA indices and language that were at least moderate in magnitude (i.e., $r \geq .30$) with a two-tailed test and $\propto = .05$.

Results

Stability

Stability differed by variable and group status (see Table 2). All variables reached acceptable stability in the Sibs-autism group within two to three observations (i.e., audio recorded samples), while most variables would require six or more observations in the Sibs-NA group to achieve acceptable stability.

AVA raw scores surpassed the a priori threshold of $g \geq .8$ with two recordings as derived across both groups (g for two recordings = 0.869) and within the Sibs-autism group (g for two recordings = 0.884; See Figure 1A). This variable surpassed the threshold for acceptable stability in the Sibs-NA group after three recording days ($g = .857$).

RVC scores were stable across groups within two days of recording ($g = .830$; see Figure 1B). This index surpassed the stability threshold with two recordings in the Sibs-autism group ($g = .883$). However, analyses indicated that it would take six days of recording to obtain acceptably stable RVC scores in the Sibs-NA group ($g = .820$).

Child Vocalization Count and Child Vocalization Duration were not stable with two observations across groups (gs = .719 and .663, respectively; see Figures 1C and 1D). Both variables required a minimum of four recordings to achieve adequate stability across groups (gs = .837 and .797 for Child Vocalization Count and Child Vocalization Duration, respectively). In the Sibs-autism group, Child Vocalization Count reached acceptable stability within two recordings ($g = .797$), and Child Vocalization Duration would require a minimum of three recordings to reach acceptable stability ($g = .818$). In the Sibs-NA group, Child Vocalization Count would reach acceptable stability within six recordings ($g = .814$), while Child Vocalization Duration would not reach sufficient stability even with six recordings ($g = .745$).

IVD scores were stable across groups after three recording sessions (g = .825) and within the Sibs-autism group after two recording sessions (g = .848) but would not surpass the a priori threshold even with six recordings within the Sibs-NA group (g = .568, see Figure 1E).

ACPU-C+V scores were sufficiently stable after two recording sessions both across groups (g = .918) and within groups (g = .918 and .920 for two recordings in Sibs-autism and Sibs-NA groups, respectively; see Figure 1F). ACPU-C+V scores and AVA raw scores were the only variables that demonstrated adequate stability in the Sibs-NA infant group in fewer than six observations (g = .885 and .800 respectively, after two recording sessions).

Concurrent Validity

Expressive Language

Like stability, validity differed by variable and group status (see Table 3). Despite demonstrating relatively high stability, neither AVA raw scores (zero-order correlation = 0.18) nor RVC scores (zero-order correlation = 0.14) were associated with concurrent expressive language across groups. Across groups, IVD scores were significantly negatively associated with expressive language (zero-order correlation = –0.37, p = 0.041), such that scores indexing less complex vocalizations tended to be associated with greater concurrent language abilities. The association with ACPU-C+V scores and concurrent expressive language was not significant across groups (zero-order correlation = –0.20). The aforementioned findings for IVD and ACPU-C+V scores were, notably, not in the anticipated direction.

Multiple regression analyses indicated that the relations between the expressive language aggregate and Child Vocalization Count and Duration were moderated by group (p values for vocal index*likelihood group product terms in multiple regression models testing moderated effects = 0.066 and 0.021, respectively; see Figure 2). Within the Sibs-NA group, Child

Vocalization Count (zero-order correlation = 0.35) and Child Vocalization Duration (zero-order correlation = 0.40) were moderately and positively associated with concurrent expressive language. Contrary to our expectations, Child Vocalization Count and Duration were not associated with concurrent expressive language and, in fact, trended in the opposite direction in the Sibs-autism group (zero-order correlations = –0.19 and –0.29, respectively).

Receptive Language

Neither Child Vocalization Count nor Child Vocalization Duration was significantly associated with concurrent receptive language across groups, despite small to moderate effect sizes (zero-order correlations = 0.30 and .26 for associations between Child Vocalization Count and Child Vocalization Duration and receptive language, respectively). IVD scores were also not significantly associated with receptive language across groups (zero-order correlation = –0.27). The associations for the ACPU-C+V score with concurrent receptive language was not significant across groups (zero-order correlation = –0.07). Similar to relations with expressive language, the associations with the IVD and ACPU-C+V scores and concurrent receptive language were not trending in the anticipated direction. Product terms testing moderated effects were non-significant for these models (see Table 2).

Concurrent Validity Follow-Up Analyses

Given the unanticipated finding of associations in the opposite of the theoretically supported direction in probes of significant moderated relations, zero-order correlations between language aggregates and the remaining LENA variables were also derived within the sibling groups (see Table 3 for a detailed summary of all correlations within and across groups). For ACPU-C+V scores, the correlations tended to be more positive in the Sibs-NA group compared to the Sibs-autism group, while correlations for RVC scores tended to be more positive in the

Sibs-autism group compared to the Sibs-NA group. The IVD score was significantly, but negatively associated with expressive language within the Sibs-autism group (zero-order correlations = –0.46). Note again that this result reflects a moderate association that is not in the anticipated direction. None of the other associations surpassed the threshold for statistical significance within groups.

Predictive Validity

Expressive Language

Across groups, Child Vocalization Count and Duration demonstrated significant positive associations with Time 2 expressive language (zero-order correlations = 0.38 and 0.38 respectively; see Figure 2), with moderate effect sizes (see Table 4). No other automated vocal indices were significantly associated with Time 2 expressive language across groups.

The association between Child Vocalization Duration and expressive language was moderated by group (*p* value for vocal index*likelihood group product terms in multiple regression models testing moderated effects = 0.089). Although the concurrent association between Child Vocalization Count and expressive language was moderated by group, the predictive association did not cross the threshold for statistical significance (*p* value for vocal index*likelihood group product term in multiple regression model testing moderated effects = 0.134; see Figure 3). For both Child Vocalization Count and Child Vocalization Duration, however, the relations were more positive in the Sibs-NA group (zero-order correlations = 0.46 and 0.50 for Child Vocalization Count and Child Vocalization Duration, respectively) compared to the Sibs-autism group (zero-order correlations = 0.26 and 0.30 for Child Vocalization Count and Child Vocalization Duration, respectively). Associations between expressive language and the remaining automated vocal indices were not moderated by group.

Receptive Language

Across groups, Child Vocalization Count was the only vocal variable significantly associated with Time 2 receptive language (zero-order correlation = 0.32, p = 0.046). No other significant associations were found between any of the automated indices of vocal development and Time 2 receptive language across groups; no associations with receptive language were moderated by group (see Table 4).

SES as a Covariate in Analyses

Maternal education level, as a proxy for SES, was significantly associated with two of the six LENA indices (i.e., CVC and AVA scores; r = .34 and .43, p = .029 and .006, respectively). This putative index of SES was not, however, associated with concurrent or future language (rs < .19, ps > .23 for indices of language at both timepoints).

Results for significant relations between LENA indices and language were robust to covarying for maternal education level. Notably, though, the association between IVD and Time 1 expressive language was significantly moderated by group when maternal education was included as a covariate, and the association between child vocalization duration and Time 2 expressive language was not significantly moderated by group when the covariate was included. We have reported associations between LENA indices and language, covarying for maternal education, in Supplemental Tables S1 and S2.

Post-Hoc Analyses

To assess whether group differences in variance on language scores could potentially account for the differential findings regarding the concurrent and predictive validity of LENA indices by sibling group, Levene's tests for equality of variances were run on the receptive and expressive language aggregates from Times 1 and 2. Sibs-autism and Sibs-NA did significantly differ in their variance of receptive and expressive language scores at Time 1 (p = .05 and .01 for receptive and expressive scores, respectively), such that Sibs-NA presented with increased variance in language to explain at this measurement period relative to Sibs-NA. There were no significant differences in the variance of receptive and expressive language scores between sibling groups at Time 2 (p = .17 and .09, respectively), though between-group differences trended in the same direction as Time 1 at this later timepoint.

Summary

In summary, stability varied by LENA variable and sibling group, such that Sibs-autism overall tended to display higher stability for automated indices of vocal development when compared to Sibs-NA. LENA variables showed limited associations with concurrent language, such that only associations with IVD were significant. Notably, this association was in the unexpected direction, such that IVD was negatively associated with concurrent language across all infants. There was more support for some LENA variables demonstrating associations with later language (i.e., predictive validity), such that CVC was positively associated with later child expressive and receptive language and CVD was positively associated with later expressive language; the latter association (i.e., between CVD and expressive language) was significantly stronger in Sibs-NA versus Sibs-autism.

Discussion

The present study evaluated the stability and validity of several indices that can be derived via automated vocal analysis in infants at high and relatively lower likelihood for a future diagnosis of autism and language impairment. To our knowledge, this is the first study to examine the stability of LENA variables in Sibs-autism and extends prior work by examining the concurrent and predictive validity of a wide range of LENA variables in Sibs-autism (Seidl et al., 2018; Swanson et al., 2018). Results of the present study suggest that it is feasible to obtain acceptably stable scores for many automated indices of vocal development that can be obtained from daylong LENA recordings, at least in Sibs-autism. Doing so would require only two days of audio recording in everyday settings for five of the six automated indices of vocal development assessed. Stability was notably poorer in Sibs-NA compared to Sibs-autism, with four of the six automated vocal indices assessed requiring six or more observations to reach sufficient stability in the Sibs-NA group. Many of the LENA variables did not show strong associations with concurrent and future child language across sibling groups.

Findings for Stability According to Sibling Group

It is unclear why stability differed according to group. We hypothesize this result may be explained by the fact that Sibs-autism tended to present with greater variability on automated indices on the whole relative to Sibs-NA, resulting in relatively more consistent rankings and thus higher stability metrics within the former versus the latter group. Alternatively, differential stability may be due to the fact that infants in the Sibs-autism group may be exposed to more consistent caregiver-maintained supported joint engagement, which could subsequently elicit child vocalizations with more reliable features (Bottema-Beutel et al., 2019). Another potential explanation for differing group stability is that the home language environment of Sibs-NA may be less stable as compared to Sibs-autism; Sibs-autism may have more consistent interactions in their home settings due to their older autistic sibling being enrolled in therapies and possibly having more structured household routines. It should be noted that high stability for many automated metrics was observed (at least in Sibs-autism) in spite of what have been considered suboptimal speech classification algorithms employed by the LENA recordings (Cristia et al., 2020), a potential source of measurement error that could certainly influence stability of scores derived via automated vocal analysis.

Concurrent and Predictive Validity of LENA Variables

Though stability of LENA variables was generally poor in Sibs-NA, the Child Vocalization Count and Duration variables demonstrated moderate to strong concurrent and predictive associations with expressive language in this group. It is notable that these observed correlations likely underestimate the true magnitude of these associations due to the instability of these particular automated scores in the relatively low likelihood infants (Cronbach et al., 1963). These specific automated metrics, thus, appear to show the greatest promise for measurement of early vocal characteristics, at least in Sibs-NA (note that small to moderate, albeit not statistically significant predictive associations appeared to be present for these variables in Sibs-autism as well). Those considering employing such indices should consider, however, that to obtain highly stable estimates of child vocalization frequency and duration via automated vocal analyses in some groups a large number of audio recordings (i.e., 6+ for Sibs-NA) would need to be collected, which may or may not be feasible in research and/or clinical practice settings.

Findings on the whole indicate that the automated scores we tested here may have limited validity or clinical utility for indexing language development in Sibs-autism, at least in the 12-18-month window, despite relatively higher stability in this group. As indicated above, the variables with seemingly the most empirical support in the present study are Child Vocalization Count and Duration, as these variables display some theorized associations with later expressive, and to a lesser degree later receptive, language across and within groups. Associations even for these variables, however, tended to be attenuated in Sibs-autism in comparison to Sibs-NA. It is possible that the limited variability in language scores at Time 1 in the Sibs-autism relative to the

Sibs-NA accounts, at least in part, for more attenuated associations between LENA indices and language outcomes in the former versus the latter group.

It is notable that several of the variables that had previously amassed some psychometric support for use with preschool aged autistic children, including RVC, IVD, and ACPU-C+V scores, were not correlated with concurrent or later language across groups or within Sibs-autism, despite displaying acceptable stability in some cases (e.g., Harbison et al., 2018; Woynaroski et al., 2017; Yoder et al., 2013). In fact, of the aforementioned scores, only the ACPU-C+V score yielded relations with language that trended in the expected direction. Correlations between IVD scores and language aggregates tended to be negative (i.e., not in the anticipated direction), and associations between RVC scores and language were, on the whole, negligible in magnitude. These results suggest that additional work is much needed to ascertain the degree to which automated vocal analysis is valid for use early in life, in particular in infants at elevated likelihood for autism and other language and language impairments.

Using LENA in Sibs-autism

The limited validity of selected LENA variables in Sibs-autism (despite relatively higher stability as compared to Sibs-NA) may be due to the inability of automated vocal analysis to tap aspects of vocal development that may be particularly important for this population. For example, at present, there is not an automated LENA variable that differentiates between vocalizations that are communicative verses noncommunicative in nature. This distinction may be particularly important, as toddlers who go on to receive an autism diagnosis have been shown to vocalize more frequently for noncommunicative rather than communicative purposes relative to their typically developing peers, and communicative vocalizations have been observed to be more strongly associated with future language outcomes than noncommunicative vocalizations (e.g., Plumb & Wetherby, 2013; Shumway & Wetherby, 2009). Future studies may, therefore, further explore the extent to which conventionally coded indices of vocal development appear to be more valid for use in Sibs-autism; some work has already been done investigating the benefits of human annotation and coding in addition to automated methods of vocal development in autistic children and in Sibs-autism (e.g., Edmunds, 2019; McDaniel et al., 2020).

Limitations and Future Directions

This study provides new insights into the psychometrics of automated vocal analysis in infants at high and relatively lower likelihood for autism but is not without limitations. The greatest of these limitations is that the correlational design employed does not control for alternative explanations for relations of interest. Thus, additional work is needed before we can draw conclusions regarding the causal nature of any of the (limited) associations observed here.

Two limitations must be noted about our convenience sample. First, the sample was much more homogenous than we had hoped in regards to race, ethnicity, and SES. Second, we did not collect information regarding the participants' home dialects given that (a) the infants in our sample were too young to formally assess dialect via previously developed and validated screening tools and (b) there is, to our knowledge, no standardized and/or norm-referenced assessment for caregivers' home dialect use. Thus, our results may not generalize to the full population of autistic children, as these dimensions of diversity (i.e., race, ethnicity, socioeconomic status, dialect) may interact to influence early language development in different ways (Diemer et al., 2013; McLaughlin et al., 2021; Oetting, 2020). Future work should investigate these concurrent and predictive associations in groups of Sibs-autism who are more diverse.

Additionally, the small sample size and preliminary nature of this study may have limited our ability to detect effects of interest, as several correlations and moderated effects were smaller in magnitude than we had anticipated. A priori power analyses indicated that the present study was powered to detect moderate to large correlations between LENA indices and language, but not associations of lesser magnitude. Additional work involving larger samples is, therefore, required to more fully evaluate the validity of automated vocal analysis in Sibs-autism. For example, completing an individual participant-level data meta-analysis could provide more precise estimates regarding the magnitude and directionality of concurrent and predictive associations of LENA variables in infants at increased and general-population level likelihood for a future diagnosis of autism.

Finally, future work should investigate associations between LENA indices and language outcomes in Sibs-autism at later timepoints. We measured language outcomes in the 21-27-month window, which is before autism diagnoses are considered stable in this population (Ozonoff et al., 2015). Following this sample to later timepoints will allow us to draw conclusions about which automated indices of vocal development most strongly predict language impairment, as well as autism, in Sibs-autism, who are at increased likelihood for these diagnoses.

Conclusion

The findings of this preliminary study contribute to our understanding of the stability and utility of using LENA to measure prelinguistic vocal development in infants, including those who are at increased familial likelihood for autism. Our results suggest that the stability of LENA variables differs between infants at increased likelihood for autism and infants at a lower, population-level likelihood for autism such that infants at increased likelihood for autism demonstrated higher stability on automated LENA variables. Additionally, few LENA variables that were examined in this study had validity for predicting concurrent and future language in these populations, although there was slightly more empirical support for predicting concurrent and future language via LENA variables in infants with non-autistic siblings within the developmental windows of interest to the present report. Future work is needed to understand why these group differences may exist, and to further validate the use of automated vocal analyses to examine early language development in infants.

Acknowledgements

This work was supported by NIH/NICHD P50HD103537 (PI: Neul), NIH/NCATS KL2TR000446 (PI: Woynaroski), NIH/NIDCD 1R21DC016144 (PI: Woynaroski), NIH/NIDCD R01DC020186 (PI: Woynaroski), NIH/NIDCD F31DC020129 (PI: Dunham), NIH/NCATS TL1TR002244 (PI: Hartmann), COVID Recovery Awards from Vanderbilt Edge for Scholars and the Vanderbilt Kennedy Center (Nicholas Hobbs Award; PI: Woynaroski), an Undergraduate Summer Research Award from the Autism Science Foundation (awardee: Suzman), and the Vanderbilt Institute for Clinical and Translation Research awards VR14759 (PIs: Cascio & Woynaroski), VR21082 (PI: Santapuram), VR51537 (PI: Santapuram), and VR53067 (PI: Daly). The authors are grateful to the Adventure Science Center in Nashville, TN for their Living Labs Partnership.

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on request.

References

Bottema-Beutel, K., Kim, S. Y., Crowley, S., Augustine, A., Keceli-Kaysili, B., Feldman, J. I., & Woynaroski, T. G. (2019). The stability of joint engagement states in infant siblings of children with and without ASD: Implications for measurement practices. *Autism Research, 12*, 495-504. https://doi.org/10.1002/aur.2068

Bredin-Oja, S. L., Fielding, H., Fleming, K. K., & Warren, S. F. (2018). Clinician vs. machine: Estimating vocalizations rates in young children with developmental disorders. *American Journal of Speech-Language Pathology, 27*(3), 1066-1072. https://doi.org/10.1044/2018_AJSLP-17-0016

Chawarska, K., Paul, R., Klin, A., Hannigen, S., Dichtel, L. E., & Volkmar, F. (2007). Parental recognition of developmental problems in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*(1), 62-72. https://doi.org/10.1007/s10803-006-0330-8

Chericoni, N., de Brito Wanderley, D., Costanzo, V., Diniz-Gonçalves, A., Leitgel Gille, M., Parlato, E., Cohen, D., Apicella, F., Calderoni, S., & Muratori, F. (2016). Pre-linguistic vocal trajectories at 6-18 months of age as early markers of autism. *Frontiers in Psychology, 7*, Article 1595. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01595

Cristia, A., Bulgarelli, F., & Bergelson, E. (2020). Accuracy of the language environment analysis system segmentation and metrics: A systematic review. *Journal of Speech, Language, and Hearing Research, 63*(4), 1093-1105. https://doi.org/10.1044/2020_JSLHR-19-00017

Crocker, L., & Algina, J (1986). *Introduction to classical and modern test theory*. Harcourt.

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology, 16*(2), 137-163. https://doi.org/10.1111/j.2044-8317.1963.tb00206.x

Diemer, M. A., Mistry, R. S., Wadsworth, M. E., López, I., & Reimers, F. (2013). Best practices in conceptualizing and measuring social class in psychological research. *Analyses of Social Issues and Public Policy, 13*(1), 77–113. http://doi.org/10.1111/asap.12001

Dykstra, J. R., Sabatos-DeVito, M. G., Irvin, D. W., Boyd, B. A., Hume, K. A., & Odom, S. L. (2013). Using the Language Environment Analysis (LENA) system in preschool classrooms with children with autism spectrum disorders. *Autism*, *17*(5), 582–594. https://doi.org/10.1177/1362361312446206

Edmunds, S. R. (2019). *The Role of Parent Responsiveness and Vocal Complexity in Expressive Language Outcome: Intervention Effects in Toddlers at Risk for ASD and Language Impairment* (Publication No. 22618443) [Doctoral dissertation, University of Washington]. ProQuest Dissertations and Theses Global.

Feldman, J. I., Raj, S., Santapuram, P., Golden, A. J., Daly, C., Dunham, K., Suzman, E., Augustine, A. E., Garla, V., Muhumuza, A., Cascio, C. J., Williams, K. L., Kirby, A. V., Keceli-Kaysili, B., & Woynaroski, T. G. (2021). Sensory responsiveness is linked with communication in infants siblings of children with and without autism. *Journal of Speech, Language, and Hearing Research, 64*(6), 1964-1976. https://doi.org/10.1044/2021_JSLHR-20-00196

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., & Reznick, J. S. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual*. Brookes.

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Oller, D. K., Hansen, J. H. L., & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology, 26*(2), 248-265. https://doi.org/10.1044/2016_AJSLP-15-0169

Gilkerson, J., & Richards, J. (2020). *A Guide to Understanding the Design and Purpose of the LENA System* (Technical Report LTR-12; LENA Technical Report). LENA Foundation. https://www.lena.org/wp-content/uploads/2020/07/LTR-12_How_LENA_Works.pdf

Graham, S., Hebert, M., Paige Sandbank, M., & Harris, K. R. (2016). Assessing the writing achievement of young struggling writers: Application of generalizability theory. *Learning Disability Quarterly*, *39*(2), 72–82. https://doi.org/10.1177/0731948714555019

Harbison, A. L., Woynaroski, T., Tapp, J., Wade, J. W., Warlaumont, A. S., & Yoder, P. J. (2018). A new measure of child vocal reciprocity in children with autism spectrum disorder. *Autism Research, 11*, 903-915. https://doi.org/10.1002/aur.1942

Hoff, E., Burridge, A., Ribot, K. M., & Giguere, D. (2018). Language specificity in the relation of maternal education to bilingual children's vocabulary growth. *Developmental Psychology, 54*(6), 1011–1019. https://doi.org/10.1037/dev0000492

Huttenlocher J, Vasilyeva M, Cymerman E, & Levine S (2002). Language input and child syntax. *Cognitive Psychology, 45*, 337–374. https://doi.org/10.1016/s0010-0285(02)00500-5

Jensen, T. S., Bøggild-Andersen, B., Schmidt, J., Ankerhus, J., & Hansen, E. (1988). Perinatal risk factors and first-year vocalizations: Influence on preschool language and motor performance. *Developmental Medicine & Child Neurology, 30*(2), 153-161. https://doi.org/10.1111/j.1469-8749.1988.tb04746.x

Jones, R. M., Plesa Skwerer, D., Pawar, R., Hamo, A., Carberry, C., Ajodan, E. L., Caulley, D., Silverman, M. R., McAdoo, S., Meyer, S., Yoder, A., Clements, M., Lord, C., & Tager-Flusberg, H. (2019). How effective is LENA in detecting speech vocalizations and language produced by children and adolescents with ASD in different contexts? *Autism Research*, *12*(4), 628–635. https://doi.org/10.1002/aur.2071

Justice, L. M., Jiang, H., Bates, R., & Koury, A. (2020). Language disparities related to maternal education emerge by two years in a low-income sample. *Maternal and Child Health Journal, 24*(11), 1419-1427. https://doi.org/10.1007/s10995-020-02973-9

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism Diagnostic Observation Schedule, second edition (ADOS-2) manual (Part I): Modules 1-4*. Western Psychological Services.

Markfeld, J. E., Feldman, J., Bordman, S., Daly, C., Santapuram, P., Humphreys, K. L., Keceli-Kaysili, B., & Woynaroski, T. Associations between caregiver stress and language outcomes in infants with autistic and non-autistic siblings. *Journal of Speech, Language, and Hearing Research.* Advance online publication. https://doi.org/10.1044/2022_JSLHR-22-00154

McDaniel, J., D'Ambrose Slaboch, K., & Yoder, P. (2018). A meta-analysis of the association between vocalizations and expressive language in children with autism spectrum disorder. *Research in Developmental Disabilities, 72*, 202-213. https://doi.org/10.1016/j.ridd.2017.11.010

McDaniel, J., Woynaroski, T. G., Keceli-Kaysili, B., Watson, L. R., & Yoder, P. (2019). Vocal communication with canonical syllables predicts later expressive language skills in preschool-aged children with autism spectrum disorder. *Journal of Speech, Language,* *and Hearing Research, 62*(10), 3826-3833. https://doi.org/10.1044/2019_JSLHR-L-19-0162

McDaniel, J., Yoder, P. J., Estes, A., & Rogers, S. J. (2020). Predicting expressive language from early vocalizations in young children with autism spectrum disorder: Which vocal measure is best? *Journal of Speech, Language, and Hearing Research, 63*(5), 1509-1520. https://doi.org/10.1044/2020_JSLHR-19-00281

McLaughlin, K. A., Sheridan, M. A., Humphreys, K. L., Belsky, J., & Ellis, B. J. (2021). The value of dimensional models of early experience: Thinking clearly about concepts and categories. *Perspectives on Psychological Science, 16*(6), 1463–1472. https://doi.org/10.1177/1745691621992346

Messinger, D., Young, G. S., Ozonoff, S., Dobkins, K., Carter, A., Zwaigenbaum, L., Landa, R. J., Charman, T., Stone, W. L., Constantino, J. N., Hutman, T., Carver, L. J., Bryson, S., Iverson, J. M., Strauss, M. S., Rogers, S. J., & Sigman, M. (2013). Beyond autism: A baby siblings research consortium study of high-risk children at three years of age. *Journal of the American Academy of Child & Adolescent Psychiatry, 52*(3), 300-308. https://doi.org/10.1016/j.jaac.2012.12.011

Mullen, E. M. (1995). *Mullen Scales of Early Learning*. American Guidance Service.

Oetting, J. B. (2020). From my perspective/opinion: General American English as a dialect: A call for change. *The ASHA LeaderLive*. https://doi.org/10.1044/leader.fmp.25112020.12

Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press. https://doi.org/10.4324/9781410602565

Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., Yapanel, U., & Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences, 107*(30), 13354-13359. https://doi.org/10.1073/pnas.1003882107

Ozonoff, S., Young, G. S., Belding, A., Hill, M., Hill, A., Hutman, T., Johnson, S., Miller, M., Rogers, S. J., Schwichtenberg, A. J., Steinfeld, M., & Iosif, A.-M. (2014, 2014/04/01/). The broader autism phenotype in infancy: When does it emerge? *Journal of the American Academy of Child & Adolescent Psychiatry, 53*(4), 398-407. https://doi.org/10.1016/j.jaac.2013.12.020

Patten, E., Belardi, K., Baranek, G. T., Watson, L. R., Labban, J. D., & Oller, D. K. (2014). Vocal patterns in infants with autism spectrum disorder: Canonical babbling status and vocalization frequency. *Journal of Autism and Developmental Disorders, 44*(10), 2413-2428. https://doi.org/10.1007/s10803-014-2047-4

Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., & Klin, A. (2011). Out of the mouths of babes: vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry, 52*(5), 588-598. https://doi.org/10.1111/j.1469-7610.2010.02332.x

Plumb, A. M., & Wetherby, A. M. (2013). Vocalization development in toddlers with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research, 56*(2), 721-734. https://doi.org/10.1044/1092-4388(2012/11-0104)

Putnam, O. C., Markfeld, J. E., Goldblum, J., Wright, S., Swanson, M. R., & Harrop, C. (2023). *The use of LENA in autism research: A systematic review* [Manuscript in preparation].

R Core Team. (2020). *R: A language and environment for statistical computing.* In (Version 4.0.2) Vienna, Austria. https://www.R-project.org/

Rankine, J., Li, E., Lurie, S., Rieger, H., Fourie, E., Siper, P. M., Wang, A. T., Buxbaum, J. D., & Kolevzon, A. (2017). Language ENvironment Analysis (LENA) in Phelan-McDermid syndrome: Validity and suggestions for use in minimally verbal children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(6), 1605-1617. https://doi.org/10.1007/s10803-017-3082-8

Richards, J. A., Xu, D., Gilkerson, J., Yapanel, U., Gray, S., & Paul, T. (2017). Automated assessment of child vocalization development using LENA. *Journal of Speech, Language, and Hearing Research, 60*(7), 2047-2063. https://doi.org/10.1044/2017_JSLHR-L-16-0157

Rogers, S. J., Yoder, P., Estes, A., Warren, Z., McEachin, J., Munson, J., Rocha, M., Greenson, J., Wallace, L., Gardner, E., Dawson, G., Sugar, C. A., Hellemann, G., & Whelan, F. (2021). A multisite randomized controlled trial comparing the effects of intervention intensity and intervention style on outcomes for young children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *60*(6), 710-722. https://doi.org/10.1016/j.jaac.2020.06.013

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin, 94*, 18-38. https://doi.org/10.1037/0033-2909.94.1.18

Rutter, M., Bailey, A., & Lord, C. (2003). *The Social Communication Questionnaire*. Western Psychological Services.

Sandbank, M., & Yoder, P. (2014). Measuring representative communication in young children with developmental delay. *Topics in Early Childhood Special Education, 34*, 133-141. https://doi.org/10.1177/0271121414528052

Santapuram, P., Feldman, J. I., Bowman, S. M., Raj, S., Suzman, E., Crowley, S., Kim, S. Y., Keceli-Kaysili, B., Bottema-Beutel, K., Lewkowicz, D. J., Wallace, M. T., &

Woynaroski, T. G. (2022). Mechanisms by which early eye gaze to the mouth during multisensory speech influences expressive communication development in infant siblings of children with and without autism. *Mind, Brain, and Education, 16*(1), 62-74. https://doi.org/10.1111/mbe.12310

Sheinkopf, S. J., Mundy, P., Oller, D. K., & Steffens, M. (2000). Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders, 30*(4), 345-354. https://doi.org/10.1023/A:1005531501155

Shumway, S., & Wetherby, A. (2009). Communicative acts of children with autism spectrum disorders in the second year of life. *Journal of Speech Language and Hearing Research, 52*(5), 1139-1156. https://doi.org/10.1044/1092-4388(2009/07-0280)1092-4388_2009_07-0280

Sparrow, S. S., Cicchetti, D. V., & Bella, D. A. (2005). *Vineland Adaptive Behavior Scales* (2nd ed.). Pearson.

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*, 112-118. https://doi.org/10.1093/bioinformatics/btr597

Sulek, R., Smith, J., Bent, C. A., Hudry, K., Trembath, D., Vivanti, G., & Dissanayake, C. (2022). The utility of LENA as an indicator of developmental outcomes for young children with autism. *International Journal of Language & Communication Disorders, 57*(1), 103-111. https://doi.org/10.1111/1460-6984.12678

Swanson, M. R., Shen, M. D., Wolff, J. J., Boyd, B., Clements, M., Rehg, J., Elison, J. T., Paterson, S., Parish-Morris, J., Chappell, J. C., Hazlett, H. C., Emerson, R. W., Botteron, K., Pandey, J., Schultz, R. T., Dager, S. R., Zwaigenbaum, L., Estes, A. M., Piven, J., &

The IBIS Network. (2018). Naturalistic language recordings reveal “hyper-vocal” infants at high familial risk for autism. *Child Development, 89*(2), e60-e73. https://doi.org/10.1111/cdev.12777

Tager-Flusberg, H., & Kasari, C. (2013, 2013/12/01). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research, 6*(6), 468-478. https://doi.org/10.1002/aur.1329

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders: Diagnosis, development, neurobiology, and behavior* (3rd ed., Vol. 1, pp. 335-364). Wiley.

Trembath, D., Westerveld, M. F., Teppala, S., Thirumanickam, A., Sulek, R., Rose, V., Tucker, M., Paynter, J., Hetzroni, O., Keen, D., & Vivanti, G. (2019). Profiles of vocalization change in children with autism receiving early intervention. *Autism Research, 12*(5), 830-842. https://doi.org/10.1002/aur.2075

Warren, S. F., Gilkerson, J., Richards, J. A., Oller, D. K., Xu, D., Yapanel, U., & Gray, S. (2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of Autism and Developmental Disorders, 40*(5), 555-569. https://doi.org/10.1007/s10803-009-0902-5

Wetherby, A. M., Cain, D. H., Yonclas, D. G., & Walker, V. G. (1988). Analysis of intentional communication of normal children from the prelinguistic to the multiword stage. *Journal of Speech, Language, and Hearing Research, 31*(2), 240-252. https://doi.org/10.1044/jshr.3102.240

Wetherby, A. M., Prizant, B. M., & Hutchinson, T. A. (1998). Communicative, social/affective, and symbolic profiles of young children with autism and pervasive developmental disorders. *American Journal of Speech-Language Pathology, 7*(2), 79-91. https://doi.org/10.1044/1058-0360.0702.79

Woynaroski, T. G. (2014). *The stability and validity of automated vocal analysis in preschoolers with autism spectrum disorder in early stages of language development* [Doctoral dissertation, Vanderbilt University]. http://hdl.handle.net/1803/14805

Woynaroski, T. G., Oller, D. K., Keceli-Kaysili, B., Xu, D., Richards, J. A., Gilkerson, J., Gray, S., & Yoder, P. J. (2017). The stability and validity of automated vocal analysis in preverbal preschoolers with autism spectrum disorder. *Autism Research, 10*, 508–519. https://doi.org/10.1002/aur.1667

Woynaroski, T. G., Watson, L. R., Gardner, E., Newsom, C. R., Keceli-Kaysili, B., & Yoder, P. J. (2016). Early predictors of growth in diversity of key consonants used in communication in initially preverbal children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 46*(3), 1013-1024. https://doi.org/10.1007/s10803-015-2647-7

Xu, D., Richards, J. A., & Gilkerson, J. (2014). Automated analysis of child phonetic production using naturalistic recordings. *Journal of Speech, Language, and Hearing Research, 57*(5), 1638-1650. https://doi.org/10.1044/2014_JSLHR-S-13-0037

Xu, D., Yapanel, U., & Gray, S. (2008). *Reliability of the LENA language environment analysis system in young children's natural home environment* [LENA Foundation Technical Report LTR-05-2]. LENA Foundation. https://www.lena.org/wp-content/uploads/2016/07/LTR-05-2_Reliability.pdf

Xu, D., Yapanel, U., Gray, S., & Baer, C. T. (2008). *The LENA language environment analysis system: The interpretive time segments (ITS) file* [LENA Foundation Technical Report LTR-04-2]. LENA Foundation. https://www.lena.org/wp-content/uploads/2016/07/LTR-04-2_ITS_File.pdf

Yankowitz, L. D., Petrulla, V., Plate, S., Tunc, B., Guthrie, W., Meera, S. S., Tena, K., Pandey, J., Swanson, M. R., Pruett, J. R., Cola, M., Russell, A., Marrus, N., Hazlett, H. C., Botteron, K., Constantino, J. N., Dager, S. R., Estes, A., Zwaigenbaum, L., Piven, J., Schultz, R. T., Parish-Morris, J., & The IBIS Network. (2022). Infants later diagnosed with autism have lower canonical babbling ratios in the first year of life. *Molecular Autism, 13*, Article 28. https://doi.org/10.1186/s13229-022-00503-8

Yoder, P. J., Lloyd, B. P., & Symons, F. J. (2018). *Observational measurement of behavior* (2nd ed.). Brookes Publishing.

Yoder, P. J., Oller, D. K., Richards, J. A., Gray, S., & Gilkerson, J. (2013). Stability and validity of an automated measure of vocal development from day-long samples in children with and without autism spectrum disorder. *Autism Research, 6*(2), 103-107. https://doi.org/10.1002/aur.1271

Yoder, P. J., Wade, J. W., Tapp, J., & Harbison, A. L. (2016). *Contingencies from LENA data.* In [Computer software]. https://github.com/HomeBankCode/LENA_contingencies

Table 1

Participant Characteristics According to Sibling Group

	Sibs-autism (n = 20) *M* (*SD*) min-max	Sibs-NA (n = 20) *M* (*SD*) min-max
Biological Sex	14 boys, 6 girls	12 boys, 8 girls
Chronological Age in Months	14.05 (1.93) 11 – 18	14.2 (2.26) 11 – 18
Mental Age in Months*	13.16 (1.52) 10.25 – 16	14.88 (2.48) 11.5 – 19.5
	n	*n*
Race	17 White 2 Multiple 1 Black	20 White
Ethnicity	1 Hispanic/Latino 19 Not Hispanic/Latino	1 Hispanic/Latino 19 Not Hispanic/Latino
Primary Caregiver's Highest Level of Education	2 High School Diploma or GED 9 College/Technical (1-2 Yrs) 4 College/Technical (3-4 Yrs) 3 Graduate/Professional School (1-2 Yrs) 2 Graduate/Professional School (3+ Yrs)	4 College/Technical (1-2 Yrs) 7 College/Technical (3-4 Yrs) 3 Graduate/Professional School (1-2 Yrs) 6 Graduate/Professional School (3+ Yrs)

Note. Sibs-NA = Infants with typically-developing older siblings, Sibs-autism = Infants with at least one older sibling diagnosed with autism, Mental age in months = Average of visual reception, fine motor, expressive language, and receptive language age equivalency scores from the Mullen Scales of Early Learning (Mullen, 1995).
*Groups significantly differed in mental age, $p = 0.012$.

Table 2

Stability by Automated Index and Sibling Group

Variable	Overall	Sibs-autism	Sibs-NA
AVA raw score	Yes (2 obs)	Yes (2 obs)	No (3 obs)
RVC	Yes (2 obs)	Yes (2 obs)	No (6 obs)
Child Vocalization Count	No (4 obs)	Yes (2 obs)	No (6 obs)
Child Vocalization Duration	No (4 obs)	No (3 obs)	No (>6 obs)
IVD	No (3 obs)	Yes (2 obs)	No (>6 obs)
ACPU-C+V	Yes (2 obs)	Yes (2 obs)	Yes (2 obs)

Note. Sibs-NA = Infants with typically-developing older siblings, Sibs-autism = Infants with at least one older sibling diagnosed with autism, AVA = Automated Vocalization Analysis score (Richards et al., 2017), RVC = Reciprocal Vocal Contingency score (Harbison et al., 2018), IVD = Infraphonological Vocal Development score (Woynaroski et al., 2017; Yoder et al., 2013), ACPU-C+V = Average Count Per Utterance-Consonant+Vowel (Xu et al., 2014). Variables denoted with "yes" reached our a priori level of stability (i.e., $g \geq 0.8$) within two observations. The number of observations required to derive a stable estimate is also provided for ease of reference below the yes/no indicator.

Table 3

Concurrent Validity by Automated Index and Sibling Group

	Overall		Sibs-autism		Sibs-NA	
Variable	Exp Lang	Rec Lang	Exp Lang	Rec Lang	Exp Lang	Rec Lang
AVA raw score	0.18	0.09	–0.10	–0.07	0.10	–0.05
RVC	0.14	0.09	0.18	0.13	–0.17	–0.24
Child Vocalization Count	0.17†	0.30	–0.19	0.20	0.35	0.32
Child Vocalization Duration	0.12†	0.26	–0.29	0.16	0.40	0.34
IVD	–0.37*	–0.27	–0.46*	–0.27	–0.14	–0.09
ACPU-C+V	–0.01	0.03	–0.20	–0.07	0.27	0.22

Note. Sibs-NA = Infants with typically-developing older siblings, Sibs-autism = Infants with at least one older sibling diagnosed with autism, AVA = Automated Vocalization Analysis raw score (Richards et al., 2017), RVC = Reciprocal Vocal Contingency score (Harbison et al., 2018), IVD = Infraphonological Vocal Development score (Woynaroski et al., 2017; Yoder et al., 2013), ACPU-C+V = Average Count Per Utterance-Consonant+Vowel score (Woynaroski et al., 2017; Xu et al., 2014), Exp Lang = Expressive language aggregate, Rec Lang = Receptive language aggregate.

†Association was moderated by Sibling group (*p* for interaction term in multiple regression analyses < 0.1).

$^{*}p < 0.05$.

Table 4

Predictive Validity by Automated Index and Sibling Group

Variable	Overall		Sibs-autism		Sibs-NA	
	Exp Lang	Rec Lang	Exp Lang	Rec Lang	Exp Lang	Rec Lang
AVA raw score	0.13	0.15	-0.18	0.12	0.08	-0.04
RVC	0.15	0.11	0.06	0.22	0.02	-0.23
Child Vocalization Count	0.38*	0.32*	0.26	0.33	0.46*	0.29
Child Vocalization Duration	0.38[†]*	0.29	0.30	0.29	0.50*	0.30
IVD	-0.17	-0.28	0.06	-0.30	-0.11	-0.13
ACPU-C+V	0.13	0.09	0.06	0.06	0.37	0.24

Note. Sibs-NA = Infants with typically-developing older siblings, Sibs-autism = Infants with at least one older sibling diagnosed with autism, AVA = Automated Vocalization Analysis raw score (Richards et al., 2017), RVC = Reciprocal Vocal Contingency score (Harbison et al., 2018), IVD = Infraphonological Vocal Development score (Woynaroski et al., 2017; Yoder et al., 2013), ACPU-C+V = Average Count Per Utterance-Consonant+Vowel score (Woynaroski et al., 2017; Xu et al., 2014), Exp Lang = Expressive language aggregate, Rec Lang = Receptive language aggregate.

[†]Association was moderated by Sibling group (*p* for interaction term in multiple regression analyses < 0.1).

*$p < 0.05$.

Figure 1

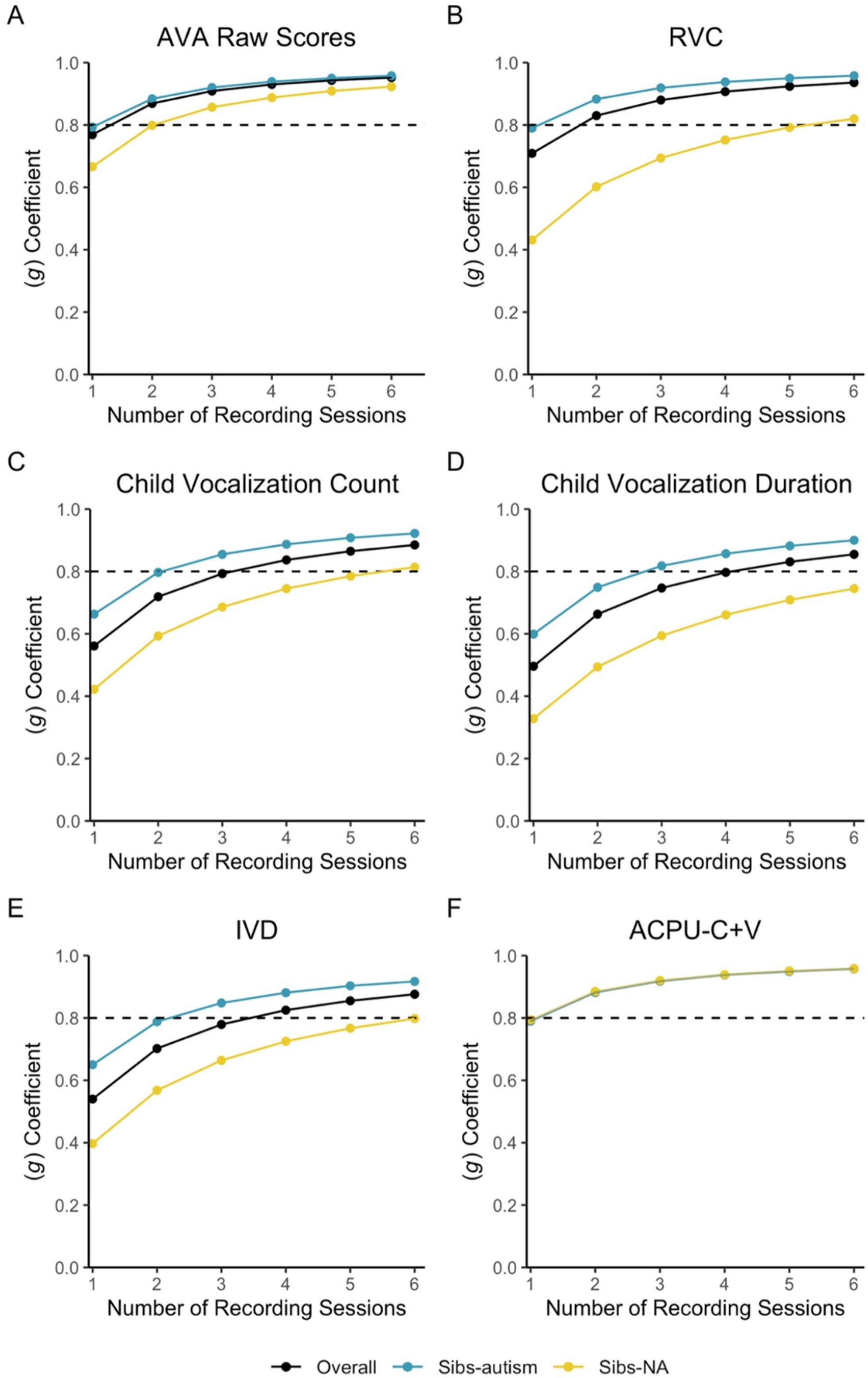
A
AVA Raw Scores
B
RVC
C
Child Vocalization Count
D
Child Vocalization Duration
E
IVD
F
ACPU-C+V
(g) Coefficient
1.0
0.8
0.6
0.4
0.2
0.0
1 2 3 4 5 6
Number of Recording Sessions
Overall
Sibs-autism
Sibs-NA

Figure 2

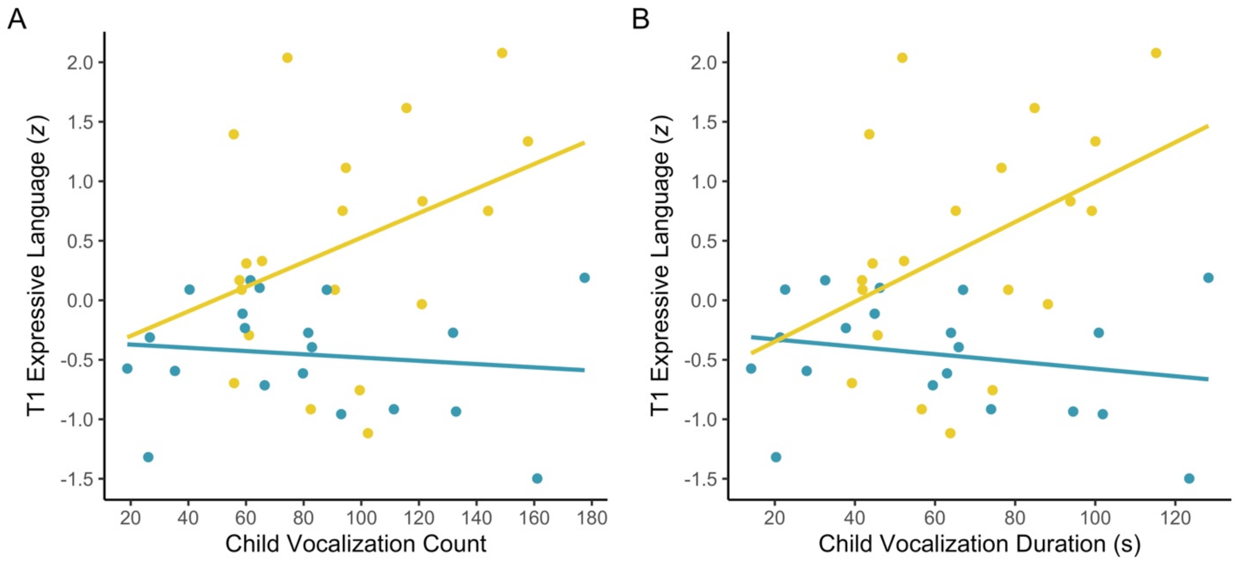
A
T1 Expressive Language (z)
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
20 40 60 80 100 120 140 160 180
Child Vocalization Count
B
T1 Expressive Language (z)
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
20 40 60 80 100 120
Child Vocalization Duration (s)

Figure 3

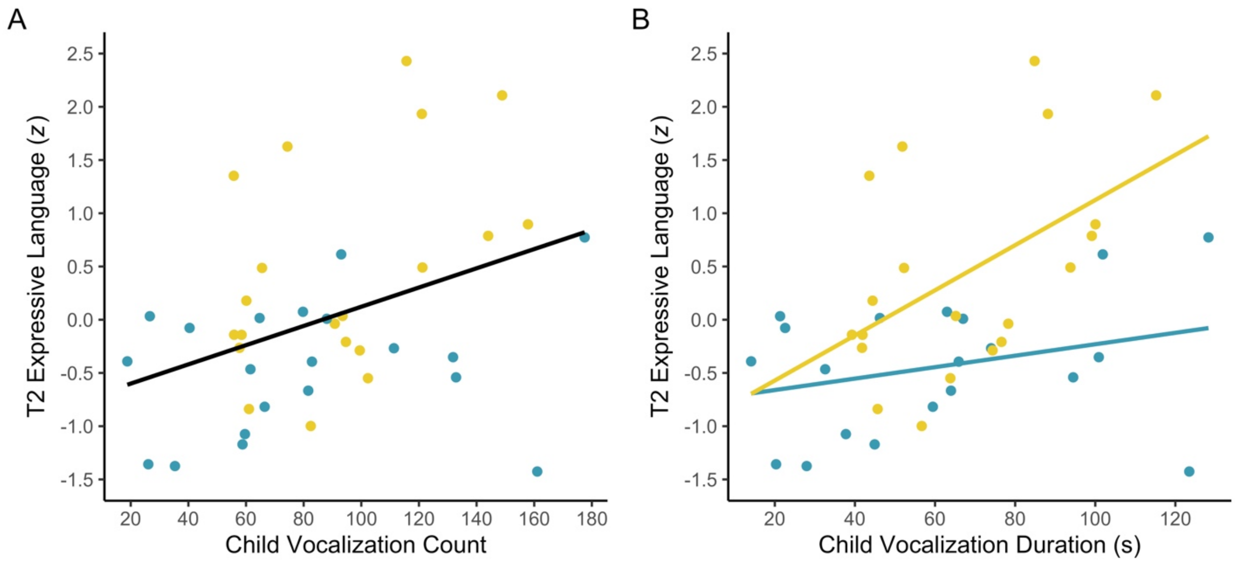
A
T2 Expressive Language (z)
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
20 40 60 80 100 120 140 160 180
Child Vocalization Count
B
T2 Expressive Language (z)
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
20 40 60 80 100 120
Child Vocalization Duration (s)

Figure Captions

Figure 1: Stability of Variables Derived from LENA Recordings. (A) Across both groups (black line) and within the Sibs-autism group (blue line), Automated Vocalization Analysis (AVA; Richards et al., 2017) raw scores reach acceptable stability ($g \geq 0.8$; dotted black line) in two observations. In the Sibs-NA group (yellow line), it would take three observations to surpass the threshold for acceptable stability. (B) Across both groups and within the Sibs-autism group, Reciprocal Vocal Contingency (RVC; Harbison et al., 2018) scores reach acceptable stability in two observations. In the Sibs-NA group, it would take six observations to reach acceptable stability. (C) Across both groups, Child Vocalization Count reaches acceptable stability in four observations. In the Sibs-autism group, it would take three observations to reach acceptable stability, and within the Sibs-NA group it would take six observations. (D) Across both groups, Child Vocalization Duration reaches acceptable stability in four observations. In the Sibs-autism group, it would take three observations to reach acceptable stability, and within the Sibs-NA group it would take over six observations. (E) Across both groups, Infraphonological Vocal Development (IVD; Yoder et al., 2013; Woynaroski et al., 2017) scores reached acceptable stability within three observations. In the Sibs-autism group, it would take two observations, and within the Sibs-NA group it would take over six observations. (F) Average Count Per Utterance - Consonant+Vowel (ACPU-C+V; Xu et al., 2014; Woynaroski et al., 2017) was the only LENA variable that was stable across and within groups within two observations. The lines are overlapping in this figure.

Figure 2: Concurrent Associations between Expressive Language and Selected LENA Variables. Concurrent associations between expressive language and (A) Child Vocalization Count and (B) Child Vocalization Duration are significantly moderated by group, such that the associations are

stronger in Sibs-NA (yellow) compared to Sibs-autism (blue). T1 = Time 1, when infants were 12-18 months of age.

Figure 3: Predictive Associations between Expressive Language and Selected LENA Variables. (A) Child Vocalization Count predicted later language across groups. (B) The association between Child Vocalization Duration and later expressive language was significantly moderated by group. The relation is significantly stronger in Sibs-NA (yellow line) compared to Sibs-autism (blue line).