Published in final edited form as:

J Autism Dev Disord. 2017 June ; 47(6): 1605–1617. doi:10.1007/s10803-017-3082-8.

Language ENvironment Analysis (LENA) in Phelan-McDermid Syndrome: Validity and Suggestions for Use in Minimally Verbal Children with Autism Spectrum Disorder

Jacquelin Rankine¹, Erin Li¹, Stacey Lurie^{1,8}, Hillary Rieger¹, Emily Fourie¹, Paige M. Siper^{1,2}, A. Ting Wang^{1,2,4,6}, Joseph D. Buxbaum^{1,2,4,5,6,7}, and Alexander Kolevzon^{1,2,3,4,5}

¹Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, Box 1668, One Gustave L. Levy Place, New York, NY 10029, USA

²Department of Psychiatry, Icahn School of Medicine at Mount Sinai, Box 1230, One Gustave L. Levy Place, New York, NY 10029, USA

³Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁴Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁵Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁶Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁷Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁸Present Address: Ferkauf Graduate School of Psychology, Yeshiva University, Bronx, NY, USA

Abstract

Phelan-McDermid syndrome (PMS) is a single-locus cause of developmental delay, autism spectrum disorder, and minimal verbal abilities. There is an urgent need to identify objective outcome measures of expressive language for use in this and other minimally verbal populations. One potential tool is an automated language processor called Language ENvironment Analysis

Alexander Kolevzon alexander.kolevzon@mssm.edu.

Author Contributions All authors contributed extensively to the work presented in this paper. JR designed the study, collected and analyzed the data, wrote the initial draft of the paper, and participated in revising the manuscript. EL, SL, HR, and EF helped with data collection and analysis. PS, TW, JB, and AK helped with designing the study. All authors assisted with manuscript development and participated in revising the manuscript.

Compliance with Ethical Standards

Conflict of interest Dr. Joseph Buxbaum, Director of the Seaver Autism Center at Mount Sinai, and the Icahn School of Medicine at Mount Sinai hold a shared patent for the use of Insulin-Like Growth Factor-1 in Phelan-McDermid syndrome. Dr. Alexander Kolevzon received research support from NIMH (R34 MH100276–01), NINDS (U54 NS092090–01), the Autism Science Foundation, the Seaver Foundation, Hoffmann-La Roche, and Neuren Pharmaceuticals.

Informed Consent Informed consent was obtained from caregivers of all individual participants included in the study.

Research Involving Animal and Human Rights All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Electronic supplementary material The online version of this article (doi:10.1007/s10803-017-3082-8) contains supplementary material, which is available to authorized users.

(LENA). LENA was used to obtain over 542 h of audio in 18 children with PMS. LENA performance was adequate in a subset of children with PMS, specifically younger children and those with fewer stereotypic vocalizations. One LENA-derived language measure, Vocalization Ratio, had improved accuracy in this sample and may represent a novel expressive language measure for use in severely affected populations.

Keywords

Phelan-McDermid syndrome; 22q13 deletion syndrome; Autism spectrum disorder; Automated vocal analysis; Language ENvironment Analysis; Minimally verbal

Introduction

Autism spectrum disorder (ASD) is a developmental disorder characterized by deficits in social communication, and restrictive and repetitive behaviors or interests (APA 2013). Abnormalities in expressive language are central to ASD and may include repetitive or stereotyped speech, echolalia, impaired pragmatics, and poor oral motor skills (Tager-Flusberg and Caronna 2007; Rapin and Dunn 1997, 2003). Early expressive language skills are the strongest predictor of positive outcomes, including increased verbal and nonverbal IQ, improved adaptive functioning, and fewer core ASD symptoms (Gillberg and Steffenburg 1987; Luyster et al. 2007; Howlin et al. 2004; Szatmari et al. 2003). Importantly, language abilities in ASD are known to improve with early detection and intervention (Rogers and Vismara 2008; Dawson 2008). Despite intervention, up to 30% of individuals with ASD remain nonverbal or minimally verbal (Tager-Flusberg et al. 2005). These minimally verbal children are at increased risk of adverse outcomes later in life, including increased social withdrawal and poorer quality of life (Anderson et al. 2011; Gerber et al. 2008). Unfortunately, minimally verbal children are often excluded from language studies due to difficulties in assessing this population with available instruments (Kasari et al. 2013). There is now an urgent need to identify objective clinical outcome measures in severely disabled populations.

Recently, genomic approaches have identified an increasing number of rare disease mutations in ASD and positioned the field to study personalized approaches to treatment. Loss of one functional copy (haploinsufficiency) of *SHANK3* on chromosome 22 through deletion or mutation occurs in 0.5% to 2% of ASD cases (Leblond et al. 2014; Betancur 2011; Gauthier et al. 2009; Moessner et al. 2007; Cooper et al. 2011; Gong et al. 2012) and results in a condition called Phelan-McDermid syndrome (PMS) (Betancur and Buxbaum 2013; Bonaglia et al. 2007; OMIM 606232). PMS is characterized by global developmental delay, ID, delayed or absent speech, hypotonia, and ASD (Soorya et al. 2013; Phelan et al. 2001; Cusmano-Ozog et al. 2007; Philippe et al. 2008). The severity of these symptoms varies among those affected, although language impairment is a consistent feature. *SHANK3* and associated pathways may be implicated in multiple forms of ASD and ID in addition to PMS, including tuberous sclerosis complex and Fragile X syndrome (Sakai et al. 2011; Darnell et al. 2011). Preclinical studies of *SHANK3* deficiency have clarified specific synaptic deficits and provided important targets for potentially disease modifying

interventions (Bozdagi et al. 2010). Drug testing in model systems of PMS has led to pilot studies in humans (Bozdagi et al. 2013; Kolevzon et al. 2014), but to further develop and rigorously evaluate treatments, there is a critical need for objective clinical outcome measures.

Traditional assessments used in minimally verbal populations have relied on parent report or structured observations to provide subjective measures of speech development and language skill (Bornstein and Haynes 1998). While these assessments provide useful information on the relative extent of speech delay, objective measures of language ability are not easily obtained. Many have questioned the utility of traditional assessments for individuals with ASD or ID who are minimally verbal and may have limited attention and motivation during formal evaluations (Koegel et al. 1997; Luyster et al. 2008). Traditional assessments are also subject to floor effects in severely language-impaired or developmentally delayed individuals and may not be sensitive to small, but clinically meaningful, changes in language abilities over time (Mervis and Robinson 2005).

An alternative technique that has been promoted in minimally verbal children with ASD or ID is the use of naturalistic language samples (Tager-Flusberg et al. 2009). This technique involves observing the child in their natural environment and collecting and transcribing language samples, usually over many hours and days. Natural language samples have long been considered the gold standard for accurately assessing expressive language abilities, but this was previously a time-consuming and labor intensive means of data collection (Hart and Risley 1995). The development of new techniques for the collection of objective data on language skills in minimally verbal individuals is critical for understanding the clinically heterogeneous phenotype of ASD and ID and for providing sensitive outcome measures for future research.

One potential tool is Language ENvironment Analysis (LENA), an automated speech analysis software that uses a digital language processor (DLP) and audio processing algorithms to record and measure language in a child's natural environment. LENA can be used to capture and automatically analyze up to 16 h of high-quality audio at home, school, therapy sessions, or anywhere a child travels in a typical day (Xu et al. 2009). After its initial validation in a population of typically developing 2–48 month old children, LENA has garnered interest from researchers and clinicians involved in the treatment of ASD and other neurodevelopmental disorders. LENA has been used to assess language skills and audio environments of children with ASD, Down syndrome, and language delay. However, LENA has never been used in a population of children with both ASD and severe language impairment.

LENA may be particularly well suited for use in this population due to its ability to capture and quantify early pre-speech vocalizations and atypical non-speech sounds that are not easily measured with traditional assessments of expressive language. The association between these early prelinguistic vocalizations and later expressive language abilities has been well established among children with typical development and developmental delay (McCathren et al. 1999; Watt et al. 2006). Although research among children with ASD is more limited, a similar association has been found. Plumb and Wetherby (2013) evaluated

50 children with ASD aged 18–24 months to determine the rate of pre-speech vocalizations and atypical non-speech vocalizations. Higher proportions of pre-speech vocalizations and lower proportions of atypical vocalizations were positively correlated with concurrent measures of speech development at 24 months and later expressive language abilities at three years of age.

The study presented here evaluates LENA in a population of minimally verbal children with ASD due to PMS for which no objective and reliable outcome measures currently exist. The primary aim is to test the validity of LENA-generated variables including speech-like vocalizations and atypical non-speech sounds as an objective measure of language ability in PMS by comparing LENA to (a) human transcription and (b) previously validated measures of expressive language ability. The secondary aim is to identify child-specific characteristics that may affect LENA reliability in order to guide further research and provide preliminary recommendations for the use of this tool in populations of minimally verbal children with ASD or ID.

Methods

Study Design

Participants were recruited as part of ongoing studies in PMS at the Seaver Autism Center for Research and Treatment at the Icahn School of Medicine at Mount Sinai. After caregivers provided informed consent, all participants underwent comprehensive multidisciplinary evaluations, including psychiatric, clinical genetics, and neurological evaluation, ASD diagnostic testing, cognitive testing with instruments selected based on age and developmental level, and an adaptive behavior assessment. All children were diagnosed with PMS using chromosomal microarray or sequencing. All study procedures were approved by the Mount Sinai Institutional Review Board.

Participants

Eighteen children participated in this study. Chronological age ranged from 30 to 172 months ($M = 76.92$, $SD = 31.78$) and was similarly distributed among the 10 female ($M = 81.22$, $SD = 38.22$) and 8 male ($M = 71.56$, $SD = 22.72$) participants. The gender distribution in the sample was representative of the larger population of PMS in which males and females are equally affected (Phelan et al. 2001).

All 18 children had major language delay as determined by scores on the Mullen Scales of Early Learning (Mullen 1995; Bishop et al. 2011) and the Vineland Adaptive Behavior Scales, Second Edition (Sparrow et al. 2005). Five children had single word speech, but none of the children used phrase speech according to parent report on the MacArthur-Bates Communicative Developmental Inventories (CDI; Fenson et al. 2007). Seventeen of the 18 children (94%) met clinical consensus criteria for ASD; all 18 children displayed ASD traits and met criteria for autism or ASD on the Autism Diagnostic Observational Schedule, Second Edition (ADOS-2) (Lord et al. 2012). Participant characteristics are summarized in Table 1.

Hardware and Software

The LENA system consists of a DLP and the associated language analysis software. The DLP records the child's vocalizations and all other sound in the child's environment within a 4- to 6-foot radius. When a recording is complete, the audio file is transferred from the DLP to a computer for automatic analysis by LENA software. The software processes the audio file using a Gaussian mixture model approach to identify and label each sound segment based on its similarity to a source model developed by LENA engineers (Gilkerson and Richards 2009). Sound segments are labeled as one of eight different types: Key Child, Adult Male, Adult Female, Other Child, Overlap, Noise, Electronic Media, and Silence. Key Child segments are further coded as true speech-related vocalizations (e.g., coos, babbles, words) or non-speech sounds (e.g., cries, yells, burps, respirations). Of note, the LENA audio processing algorithm was initially developed in typically developing children aged 2–48 months. While the chronological age of many of the children included in this study falls outside of this range, all participants had a developmental age less than 48 months old. A full report on the LENA system's automatic speech recognition technology has been previously published (Gilkerson and Richards 2008, 2009; Xu et al. 2008).

Data Collection

Each participant completed between one and three daylong recordings in their natural environment. Parents were instructed to turn on the DLP and insert it into the front pocket of the supplied LENA shirt when the child woke up in the morning. The DLP could be removed and placed nearby during bath time or naptime. Parents were instructed to choose a recording day that was representative of the child's normal routine and record for the entire day up to a maximum of 16 h. The recordings were collected in a variety of locations including at home, outdoors, at school, or at therapy sessions. At night, the parents removed the DLP and placed it in a pre-labeled envelope for shipment back to the research team. Five children completed one daylong recording, two children completed two daylong recordings, and 12 children completed three daylong recordings. A total of over 542 h of audio recording was collected from the sample.

Audio Segment Selection

For each participant, a total of three hours of audio recording was selected for human transcription. Recordings were divided into five-minute audio segments by the LENA system and segments were ordered according to the number of LENA identified child vocalizations. Thirty-six non-consecutive segments were selected for each participant, divided evenly between recording day and child vocalization count quartile. The segments selected represented a variety of activities including mealtime, bath time, story time, outdoor play, and indoor play. Segments in which no vocalizations were observed, such as naptimes, were excluded. A total of 54 h of audio recording was transcribed.

Human Transcription

A primary transcriber reviewed and coded all selected audio segments using Transcriber (Barras et al. 1998). Transcriber is a publicly available transcription software which allows the user to identify sound segment boundaries and code sound sources for various audio

formats. LENA audio files were uploaded to Transcriber and coded using a three-pass method. The transcriber first listened to the five-minute audio segment of interest and demarcated sound segment boundaries for all speakers, other noises, and periods of silence. Next, the transcriber listened to the complete audio segment again and categorized each sound segment identified in the first pass as belonging to one of the eight possible sound source categories. Finally, the transcriber completed a third pass of the audio segment and categorized each Key Child sound segment as a speech-related child vocalization or a non-speech sound. An example of sound segment identification and coding by the human transcriber is presented in Fig. 1.

A subset of 24 five-minute audio segments was randomly selected for transcription by a second rater. Both transcribers identified all Key Child sound segment boundaries and labeled each as vocalization or non-speech sound. Count estimates were highly correlated between the two transcribers for total Key Child ($r = 0.925$, $p < 0.001$), child vocalizations ($r = 0.95$, $p < 0.001$), and child non-speech sounds ($r = 0.703$, $p < 0.001$). Inter-rater reliability was excellent for the distinction of Key Child from all other segments ($k = 0.856$, $p < 0.001$), 95% CI [0.838–0.874] and the distinction of child vocalizations from non-speech sounds ($k = 0.861$, $p < 0.001$), 95% CI [0.820–0.902]. Point-by-point inter-rater reliability was calculated by Cohen's Kappa. Both transcribers were graduate students in the department of medical education who were trained in the use of Transcriber software. Human transcribers were trained to classify Key Child sound segments including identifiable consonant or vowel sounds as speech-related child vocalizations and all sound segments not meeting these criteria (e.g., laughs, yells, cries) as non-speech sounds. Transcribers were not required to perform phonetic transcription.

A total of 119,960 individual sound segments were coded across the 18 participants. Following segment identification and coding, the human transcriptions were reviewed to obtain counts of total Key Child segments, Key Child speech-related vocalizations, and Key Child non-speech sounds. Human and LENA transcriptions were compared to determine the accuracy of the LENA system for distinguishing Key Child segments from all other sounds sources, and for distinguishing Key Child vocalizations from non-speech sounds.

Measures

LENA Accuracy Measures—Two measures of LENA-human inter-rater agreement were obtained to assess the accuracy of the LENA system. Key Child inter-rater agreement was based on the coding of all 119,960 individual sound segments as Key Child or one of the other possible sound source categories. Child vocalization inter-rater agreement was based only on the 12,229 individual sound segments that both the human transcriber and LENA identified as being produced by the Key Child, in accordance with the initial LENA validation procedure (Xu et al. 2009). Child vocalization inter-rater agreement was based on the coding of these segments as true speech-related child vocalizations or non-speech sounds.

LENA Expressive Language Measures—LENA-generated measures of expressive language ability are count estimates of total Key Child segments and total child

vocalizations. Key Child and child vocalization counts were converted to rates based on total recording duration, to standardize across participants. Higher Key Child and child vocalizations rates were assumed to be potential indicators of greater expressive language abilities. LENA also calculates the total duration of Key Child vocalizations and Key Child non-speech sounds. These values were used to develop a novel language measure, the vocalization ratio, which is derived by dividing the duration of Key Child vocalizations by the duration of Key Child non-speech sounds. Higher vocalization ratio values indicate greater proportions of true speech-related vocalizations. The vocalization ratio can be easily calculated from standard LENA output and may serve as a potential novel measure of expressive language ability.

Standardized Measures—All participants were evaluated using a comprehensive assessment battery to characterize cognitive functioning, language ability, behavioral symptoms, and ASD symptom severity. These evaluations were obtained via direct assessment in the clinic and through caregiver report questionnaires. All assessments were performed by trained clinicians as part of ongoing studies (see Soorya et al. 2013; Kolevzon et al. 2014). The domains of interest assessed were as follows:

1. *ASD symptomatology* was assessed using the ADOS-2, Module 1 (Lord et al. 2012). The ADOS-2 is a semistructured observation designed to elicit social interaction, communication, and imaginative play in individuals suspected of having ASD. Module 1 is designed to assess individuals who are non-verbal or use speech at the single-word level.
2. *Expressive language* was measured with the Mullen Scales of Early Learning (Mullen 1995) and the Vineland Adaptive Behavior Scales, Second Edition, Survey Interview Form (Sparrow et al. 2005). The Mullen is a developmental assessment that provides an overall Early Learning Composite (up to 68 months) and subtest scores for developmental scales (Visual Reception, Expressive Language, Receptive Language, Gross Motor, Fine Motor). The Mullen expressive language raw score was used in this analysis. The Vineland-II Survey Interview form is a clinician-administered caregiver interview that assesses a child's skills in key domains related to adaptive functioning (Communication, Daily Living Skills, Socialization, and Motor Skills for children six years and younger). The Vineland-II expressive language raw score was used in this analysis. Age equivalent scores were not used due to methodological issues with the use of age equivalents in analysis and floor effects seen in severely delayed populations (Maloney and Larrivee 2007).
3. *Behavioral symptoms* were evaluated by the Aberrant Behavior Checklist (ABC) (Aman et al. 1985) and the Repetitive Behavior Scales-Revised (RBS-R) (Lam and Aman 2007). The ABC is a 58-question checklist that assesses problem behaviors in individuals with intellectual disability and captures symptoms in five domains: Irritability/Agitation, Stereotypic Behavior, Lethargy/Social Withdrawal, Hyperactivity/Noncompliance, and Inappropriate Speech. The ABC Inappropriate Speech subdomain raw score was used in this analysis. The RBS-R is a 44-item questionnaire designed to measure repetitive behaviors in individuals

with ASD. It consists of six subscales including: Stereotyped Behavior, Self-injurious Behavior, Compulsive Behavior, Routine Behavior, Sameness Behavior, and Restricted Behavior. Total RBS-R score was used in this analysis.

Data Analysis

LENA and human transcription inter-rater agreement was assessed for two main criteria (a) distinction of the Key Child from all other sound segment categories and; (b) distinction of Key Child vocalizations from Key Child non-speech sounds. Inter-rater agreement for each criteria was calculated with Cohen's Kappa for the entire sample and for each participant. Correlation coefficients were calculated between the two LENA accuracy measures. Correlation coefficients were also calculated for the LENA accuracy measures and potential predictors of improved performance. Potential predictors of improved LENA accuracy included in the analysis were chronological age, ASD symptom severity (i.e., ADOS-2 comparison score), and behavioral symptoms (i.e., RBS-R total score, ABC Inappropriate Speech score).

To assess the accuracy of LENA-generated expressive language measures, we reviewed the total Key Child segments and child vocalization counts generated by LENA for each child. Human transcriber count estimates were compared to LENA count estimates per participant. LENA vocalization ratio was calculated as described above from the 54 h of transcribed audio, and human transcriber vocalization ratio was calculated by dividing child vocalization count by child non-speech sounds count for each participant. Correlation coefficients were calculated to assess the consistency between LENA and human estimates.

LENA and human expressive language data were assessed for normality by the Shapiro–Wilks test. Those data found to be skewed were log-transformed and normalized before further analysis. Bland–Altman plots were constructed to visualize the level of agreement between the LENA and human estimates and to assess for any systematic differences between the two methods. The difference between paired LENA and human counts (transcriber count—LENA count) was plotted as a function of their mean ($[\text{LENA count} + \text{transcriber count}]/2$) for each potential expressive language measure. Mean differences (estimated bias) and 95% limits of agreement (LoA) were calculated. Paired sample t-tests were used to verify the absence of any systematic differences between methods. LENA expressive language measures (i.e., total Key Child rate, child vocalization rate, vocalization ratio) were compared to traditional expressive language measures (i.e., Vineland-II expressive language score, Mullen expressive language score) using correlation coefficients.

Results

LENA-Human Inter-rater Agreement: Key Child vs All Others

LENA correctly labeled 40% of the Key Child segments and misclassified them as an alternative source 60% of the time. LENA correctly labeled 97% of the segments as not being produced by the Key Child and incorrectly attributed only 3% as Key Child when they were produced by another sound source (Table 2). Inter-rater agreement between LENA and the human transcriber across the entire sample was moderate ($k = 0.449$, $p < 0.001$), 95% CI

[0.443–0.455]. Inter-rater agreement varied significantly by participant ($k = 0.144$ to $k = 0.78$) which was further examined below. The positive percent agreement (PPA) was 40.29%, 95% CI [39.73, 40.84]. The negative percent agreement (NPA) was 97.03%, 95% CI [96.92, 97.14]. Key Child counts, interrater agreement, PPA, and NPA per participant are presented in Online Resource 1.

LENA-Human Inter-rater Agreement: Key Child Vocalizations vs Non-Speech Sounds

Among the correctly labeled Key Child segments, LENA correctly labeled 84% of the child vocalizations and misclassified them as non-speech sounds 16% of the time. LENA correctly labeled 61% of the child non-speech sounds but incorrectly labeled 39% as vocalizations (Table 3). Inter-rater agreement between LENA and the human transcriber across the entire sample was moderate ($k = 0.455$, $p < 0.001$), 95% CI [0.437–0.473]. Inter-rater agreement varied significantly by participant ($k = 0.023$ to $k = 0.633$) which was further examined below. The two inter-rater agreement scores were correlated ($r = 0.514$, $p = 0.029$). The PPA was 83.82, 95% CI [83.00, 84.61]. The NPA was 60.91, 95% CI [59.40, 62.40]. Key Child vocalization counts, inter-rater agreement, PPA, and NPA per participant are presented in Online Resource 1.

Predictors of LENA Performance

As described, LENA accuracy varied widely across participants. To identify potential predictors of improved LENA performance, correlation coefficients were calculated for the LENA accuracy measures and a number of child-specific predictor variables (Table 4). Lower age was correlated with improved LENA performance for Key Child inter-rater agreement (i.e., Key Child vs All Others) ($r = -0.618$, $p = 0.006$) and child vocalization inter-rater agreement (i.e., Key Child vocalizations vs Key Child non-speech sounds) ($r = -0.577$, $p = 0.012$). Lower scores on the ABC Inappropriate Speech subdomain were correlated with improved LENA performance for child vocalization inter-rater agreement ($r = -0.567$, $p = 0.014$). ADOS-2 and RBS-R were not correlated with any LENA accuracy measures.

LENA Expressive Language Measure Accuracy

Descriptive statistics for the LENA and human transcriber estimates of Key Child count and child vocalization count are listed in Table 5. The data presented represent the average counts for the three hours of audio recording selected for human transcription for each participant. All variables were found to be significantly skewed ($p < 0.05$) with a Shapiro–Wilks test. All log-transformed variables had a normal distribution ($p > 0.05$) and are described in Table 5.

LENA and human transcriber estimates of total Key Child segments were moderately correlated ($r = 0.707$, $p = 0.001$) (Fig. 2). Across all participants, the LENA Key Child count estimate was 49% of the true human estimate. LENA estimates of the total Key Child segments ranged considerably across participants from 15 to 87% of the value estimated by the human transcriber. In all cases, LENA underestimated the number of speech segments produced by the Key Child. The Bland–Altman plot of the log-transformed data revealed a mean difference of -0.296 , 95% LoA $[-0.667, 0.075]$ (Fig. 3). There was a significant difference between the means of the two methods, $t(17) = -0.296$, $p < 0.0001$.

LENA and human transcriber estimates of child vocalizations were moderately correlated ($r = 0.60$, $p = 0.008$) (Fig. 2). Across all participants, the LENA child vocalization count estimate was 53% of the human estimate. In general, LENA underestimated the true child vocalization

Comparison of LENA and Traditional Expressive Language Measures

LENA-generated expressive language measures were compared to existing standardized expressive language measures to evaluate their validity in this population. The Vineland-II expressive language raw scores and Mullen expressive language raw scores were moderately correlated ($r = 0.635$, $p = 0.005$). Key Child rate and child vocalization rate were not correlated with either of the traditional language measures.

Evaluating Vocalization Ratio as a Novel LENA Expressive Language Measure

Descriptive statistics for the LENA and human transcriber estimates of the vocalization ratio are listed in Table 5. The variable was log-transformed and had a normal distribution ($p > 0.05$) with a Shapiro–Wilks test. LENA vocalization ratio was highly correlated with the human transcriber vocalization ratio ($r = 0.953$, $p < 0.001$) (Fig. 2). A Bland–Altman analysis was performed to assess the accuracy of this measure. The Bland–Altman plot of the logtransformed data revealed minimal bias with a mean difference of -0.038 , 95% LoA $[-0.775, 0.700]$ (Fig. 3). No significant difference between the means of the two methods was detected, $t(17) = -0.038$, $p = 0.676$, indicating that this may be an accurate measure for use in minimally verbal populations. LENA vocalization ratio was compared to traditional expressive language measures and was correlated with Vineland-II expressive language raw score ($r = 0.482$, $p = 0.043$), but was not correlated with Mullen expressive language raw score ($r = 0.306$, $p = 0.217$).

Discussion

The LENA system was assessed in a population of 18 children with PMS, a single-locus cause of ID, ASD, and minimal verbal abilities. While the LENA system has been used effectively in populations of typically developing children and children with idiopathic ASD, this study represents the first validation of LENA-generated expressive language measures for use in a population of children with ASD associated with a known genetic syndrome and minimal verbal abilities. Such a validation is important because this population may differ significantly from those with idiopathic ASD with respect to relevant variables known to influence LENA reliability, including expressive language abilities and ASD symptom severity.

A total of 648 five-minute audio segments were coded by human transcribers and compared to LENA-generated transcriptions of the corresponding data. LENA-human inter-rater agreement for the distinction of Key Child segments from all other sound sources was moderate ($k = 0.449$). The LENA system correctly identified only 40% of the Key Child segments in this sample, compared to 76% in typically developing children (Xu et al. 2009). The older chronological age of the participants in this sample made them more likely to be misclassified as Adult speech segments which may have contributed to this high false

negative rate. However, the low false positive rate for identification of the Key Child was excellent (3%) and similar to that seen in typically developing populations. LENA-human inter-rater agreement for the distinction of Key Child vocalizations from non-speech sounds was also moderate ($k = 0.455$). The LENA system had good sensitivity (84%) and fair specificity (61%) for the identification of true speech-related child vocalizations, compared to 75% sensitivity and 85% specificity in typically developing populations (Xu et al. 2009). The high false positive rate may be due in part to the presence of increased atypical nonspeech sounds (e.g., high-pitched squeals, growls) in populations of children with ASD (Schoen et al. 2011).

Correlational analyses revealed adequate reliability across the selected audio segments with respect to total Key Child ($r = 0.707$) and child vocalization ($r = 0.60$) counts. Across all participants, LENA identified 53% of child vocalization and 49% of Key Child segments. The tendency of the LENA system to underestimate these counts is in agreement with prior studies and the design of the LENA algorithm to limit false positive classifications. The improved ability of the human transcriber to identify and count vocalizations during LENA designated Overlap segments also contributed to the observed LENA underestimates of the child counts of interest. This indicates fair reliability of the LENA system for estimating total Key Child and child vocalization counts in PMS. The LENA child vocalization ratio, however, was highly correlated ($r = 0.953$) with human transcriber estimates. A Bland–Altman analysis revealed minimal bias and did not detect a significant difference the means of the LENA and human transcriber estimates of this variable.

Despite overall fair performance in this population, LENA accuracy measures varied widely across individual participants. These results suggest that LENA may provide reasonable estimates of the expressive language abilities in some minimally verbal children, but characteristics of the individual must be considered before using this tool. Recommendations may be made for the use of LENA in similar populations of minimally verbal children based on the findings of this study.

The factor most strongly correlated with LENA accuracy in this sample was participant age. LENA accuracy decreased with increasing age of the child. The LENA sound segment identification algorithm is based on a normative sample of typically developing children between 2 and 48 months of age. The sample evaluated in this study included children with chronological ages outside of this range (30 months—14 years), but no child had a verbal mental age outside of this range (3–27 months). Only two of the 18 children included in this study were under 48 months of age. Based on these findings, it is evident that LENA accuracy is dependent on chronological but not mental age. This finding is likely related to changes in the vocal acoustics of the growing child that render the LENA algorithm less reliable. Researchers and clinicians should exercise caution when using LENA as a measure of expressive language in minimally verbal children over four years of age.

It is also apparent that some features of ASD may affect the LENA-generated expressive language variables. For example, in children with high degrees of echolalia and repetitive speech, the LENA child vocalization count may be significantly elevated and less representative of meaningful speech. In this study, LENA accuracy was decreased in

participants with higher degrees of inappropriate speech as measured by the ABC. Researchers should consider scores on the ABC Inappropriate Speech subscale before using LENA in similar populations. Children with high levels of inappropriate speech, including echolalia and atypical non-speech sounds, may not be appropriate candidates for the use of LENA.

While this study suggests that LENA can provide valuable information about the expressive language abilities of children who are minimally verbal, it remains important to corroborate LENA findings through the use of multiple assessments. The LENA system only provides information about an individual's expressive language abilities. Minimally verbal individuals may engage in nonverbal communication that, while meaningful, would not be captured by the LENA system. Observed evaluations or videotaping may be used in conjunction with the LENA system to better assess nonverbal communication.

Prior studies have found significant correlations between LENA-generated and standard expressive language measures in typically developing children (Gilkerson and Richards 2009). In this study, hourly Key Child and child vocalization rates were not correlated with traditional expressive language measures. This finding does not necessarily exclude LENA as a valid measure of expressive language in this population, but may instead suggest that LENA natural language samples can provide unique information that is not captured by traditional measures.

We defined a LENA vocalization ratio by dividing the duration of Key Child vocalizations by the duration of Key Child non-speech sounds. While this variable is not reported in the LENA output, it is easily calculated from available data. The vocalization ratio was correlated with a traditional language measure, the Vineland-II expressive language raw score. The use of the vocalization ratio may serve to overcome the poor identification of Key Child segments within this population, as the ratio will remain relatively constant despite falsely low overall Key Child counts. This may represent a novel expressive language measure for use in populations of minimally verbal children with ASD who display high levels of inappropriate speech.

This study has several limitations including the study population, which was a relatively small convenience sample of participants already involved in ongoing studies. These children and their families may not be representative of the general population of children with PMS or ASD. It is also possible that language deficits in PMS are unique and findings from this population may not be generalizable to other neurodevelopmental disorders resulting in minimal verbal abilities. To improve the generalizability of these results, LENA should be evaluated in other populations including those associated with known genetic syndromes or idiopathic ASD. Gathering and transcribing more natural language samples from these children will better our understanding of heterogeneous language abilities and provide needed confirmation of the results found in this small sample.

In addition, the participant age range was wide relative to the number of participants in this sample and few children (2/18) fell within the suggested age range of the LENA device. While increasing age was clearly associated with poor LENA performance, it is difficult to

draw conclusions about a definitive age limit above which LENA is unreliable. Future evaluations of LENA reliability in populations of minimally verbal individuals should include larger samples with more children under 48 months of age. While this study showed correlations between LENA accuracy and child specific variables of interest, it will be important to define cutoff scores or composite measures of inappropriate speech, age, and other potential predictor variables above which LENA is no longer considered reliable.

Additionally, a single primary human transcriber was used due to the time intensive manner of the coding process. A secondary human transcriber was used as a check of inter-rater reliability, but coded only ~5% of the total audio evaluated by the primary transcriber. The limited number of segments transcribed may have also contributed to the discrepancies in human and LENA counts of key variables. It is important to note that the LENA system was designed to provide reliable estimates over periods of 12–16 h to account for variability across individual audio segments. In this study, only one-hour of audio data was transcribed from each daylong recording. Error rates have been found to stabilize after obtaining and transcribing at least one-hour of audio, as more individual false positive and false negative errors cancel out (Xu et al. 2009). The estimate of LENA reliability would likely continue to improve with transcription of additional audio data from each recording.

Furthermore, the samples selected for transcription were chosen to represent each quartile of child vocalization count. While this technique may have provided a better representation of the overall sample, it is reasonable to assume that segments with lower child vocalizations contribute less to the language measures obtained from the daylong recording. It is possible that the selection of segments with greater child vocalizations or the transcription of longer segment durations may have provided more accurate estimates of LENA performance in this population.

Given these limitations, we cannot make firm conclusions about the validity of individual LENA expressive language measures for use in this population. In this study, the correlation of LENA and human transcriber estimates of Key Child and child vocalizations were fair but imperfect. LENA Key Child and child vocalization counts should not be considered exact values, but instead may serve as useful indicators to rank individuals within a group or to compare performance between groups (e.g., typically developing or idiopathic ASD vs ASD associated with a known genetic syndrome). Further research is required to elucidate how LENA Key Child and child vocalization counts may be best used by researchers and clinicians as measures of expressive language ability.

It is also necessary to determine the sensitivity of LENA measures for detecting small but potentially meaningful changes in language ability over time. Future studies should include multiple time points and within participant comparisons to address this question. To better characterize how the LENA vocalization ratio can be used, further research is needed to establish test–retest stability of this measure across different recording days and settings. Longitudinal studies in which LENA-generated language measures, including the vocalization ratio, are compared to later language abilities are needed to clarify the ability of these measures to predict future language development.

Finally, no distinction was made for meaningful speech or speech directed at another individual in the coding of audio recordings. The use of the LENA-generated variables as objective measures of expressive language relies on the assumption that prelinguistic, but speech-like, child vocalizations and atypical non-speech sounds represent either current expressive language abilities or an increased potential for the development of expressive language skills. This is in line with prior work in which child vocalizations and non-speech sounds defined in the manner employed here were found to be correlated with concurrent and future expressive language abilities (Plumb and Wetherby 2013). As research on this topic in children with ASD is sparse, it is possible that these variables are not true markers of expressive language, but may better reflect a generalized behavioral tendency to use vocalizations that are speech-like. This limits the reach of this data and future studies should address the relationship between LENA child vocalization counts and meaningful speech with communicative intent.

Despite these limitations, overall, LENA accuracy in typically-developing populations is comparable to that found in a subset of this sample: younger children with lower levels of inappropriate speech and non-speech sounds. Importantly, one potential LENA expressive language measure, child vocalization ratio, had improved accuracy in this sample and was correlated with a standardized language assessment. The vocalization ratio may serve as a novel expressive language measure for use in similar populations. These results are based on 542 h of audio recording and 54 h of human transcription, representing the largest reported natural language sample of children with PMS. These findings have important implications for the use of LENA as an objective outcome measure in future trials.

The results of this study present preliminary evidence that the LENA system can provide useful information about the language output of children with ASD and severe language impairment. While the LENA system may not be reliable for all children who are minimally verbal, the findings presented here may guide the selection of appropriate candidates for the use of LENA. LENA may hold great potential as an efficient and objective tool to collect large amounts of data on the expressive language abilities of children with ASD who are minimally verbal. Future studies to examine validity and sensitivity to change over time of LENA expressive language variables in children with ASD who are minimally verbal may improve the relevance of this potentially valuable tool.

Acknowledgments

We would like to thank all the participants for taking part in our research. Jacquelin Rankine would like to thank the Icahn School of Medicine and the Patient Oriented Research Training and Leadership (PORTAL) Program for providing a research stipend during the time required to complete this work which was prepared as part of a master's thesis. We would like to thank the Icahn School of Medicine at Mount Sinai Center for Biostatistics for providing statistical support for this work. Dr. Alexander Kolevzon received research support from NIMH (R34 MH100276-01), NINDS (U54 NS092090-01), the Autism Science Foundation, the Seaver Foundation, Hoffmann-La Roche, and Neuren Pharmaceuticals.

Funding Jacquelin Rankine received a research stipend from the Icahn School of Medicine at Mount Sinai Patient Oriented Research Training and Leadership (PORTAL) Program.

References

- Aman MG, Singh NN, Stewart AW, & Field CJ (1985). The aberrant behavior checklist: a behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency*, 89, 485–491. [PubMed: 3993694]
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th edn.). Arlington, VA: American Psychiatric Publishing.
- Anderson DK, Maye MP, & Lord C (2011). Changes in maladaptive behaviors from mid-childhood to young adulthood in autism spectrum disorder. *American Journal on Intellectual and Developmental Disabilities*, 116(5), 381–397. [PubMed: 21905806]
- Barras C, Geoffrois E, Wu Z, & Liberman M (1998). Transcriber: A free tool for segmenting, labeling and transcribing speech. *First international conference on language resources and evaluation (LREC)*, 1373–1376.
- Betancur C (2011). Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. *Brain Research*, 1380, 42–77. [PubMed: 21129364]
- Betancur C, & Buxbaum JD (2013). SHANK3 haploinsufficiency: A “common” but underdiagnosed highly penetrant monogenic cause of autism spectrum disorders. *Molecular Autism*, 4(1), 17. [PubMed: 23758743]
- Bishop SL, Guthrie W, Coffing M, & Lord C (2011). Convergent validity of the Mullen scales of early learning and the differential ability scales in children with autism spectrum disorders. *American Journal on Intellectual and Developmental Disabilities*, 116(5), 331–343. [PubMed: 21905802]
- Bonaglia MC, Marelli S, Gottardi G, Zucca C, Pramparo T, Giorda R, ... Zuffardi O (2007). Subtelomeric trisomy 21q: A new benign chromosomal variant. *European Journal of Medical Genetics*, 50(1), 54–59. [PubMed: 17055792]
- Bornstein MH, & Haynes OM (1998). Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development*, 69(3), 654–671. [PubMed: 9680678]
- Bozdagi O, Sakurai T, Papapetrou D, Wang X, Dickstein DL, Takahashi N, ... Buxbaum JD (2010). Haploinsufficiency of the autism associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. *Molecular Autism*, 1(1):15. [PubMed: 21167025]
- Bozdagi O, Tavassoli T, Buxbaum JD (2013). Insulin-like growth factor-1 rescues synaptic and motor deficits in a mouse model of autism and developmental delay. *Molecular Autism*, 4(1):9. [PubMed: 23621888]
- Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, ... Eichler EE (2011). A copy number variation morbidity map of developmental delay. *Nature Genetics*, 43(9), 838–846. [PubMed: 21841781]
- Cusmano-Ozog K, Manning MA, & Hoyme HE (2007). 22q13.3 deletion syndrome: a recognizable malformation syndrome associated with marked speech and language delay. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 145(4), 393–398.
- Darnell JC, Van Driesche SJ, Zhang C, Hung KY, Mele A, Fraser CE, ... Darnell RB (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell*, 146(2), 247–261. [PubMed: 21784246]
- Dawson G (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology*, 20(03), 775–803. [PubMed: 18606031]
- Fenson L, Marchman VA, Thal DJ, Dale P, Reznick JS, Bates E (2007). *MacArthur-Bates Communicative Development Inventories: User’s guide and technical manual* (2nd ed.). Baltimore, MD: Paul H. Brookes Publishing Company.
- Gauthier J, Spiegelman D, Piton A, Lafreniere RG, Laurent S, St-Onge J, ... Rouleau GA (2009). Novel de novo SHANK3 mutation in autistic patients. *American Journal of Medical Genetics: Neuropsychiatry Genetics*, 150B(3), 421–424.
- Gerber F, Baud MA, Giroud M, & Carminati GG (2008). Quality of life of adults with pervasive developmental disorders and intellectual disabilities. *Journal of Autism and Developmental Disorders*, 38(9), 1654–1665. [PubMed: 18266098]
- Gilkerson J, Richards JA (2008). *The LENA natural language study*. Boulder, CO: LENA Foundation Retrieved March, 3, 2009.

- Gilkerson J, Richards JA (2009). The power of talk, 2nd edition. Boulder, CO: LENA Foundation Retrieved March, 3, 2009.
- Gillberg C, & Steffenburg S (1987). Outcome and prognostic factors in infantile autism and similar conditions: A populationbased study of 46 cases followed through puberty. *Journal of Autism and Developmental Disorders*, 17(2), 273–287. [PubMed: 3610999]
- Gong L, Yan Y, Xie J, Liu H, & Sun X (2012). Prediction of autism susceptibility genes based on association rules. *Journal of Neuroscience Research*, 90(6), 1119–1125. [PubMed: 22345019]
- Hart B, & Risley TR (1995). Meaningful differences in the everyday experience of young American children. Baltimore, MD:Paul H Brookes Publishing.
- Howlin P, Goode S, Hutton J, & Rutter M (2004). Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry*, 45(2), 212–229. [PubMed: 14982237]
- Kasari C, Brady N, Lord C, & Tager-Flusberg H (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research*, 6(6), 479–493. [PubMed: 24353165]
- Koegel LK, Koegel RL, & Smith A (1997). Variables related to differences in standardized test outcomes for children with autism. *Journal of Autism and Developmental Disorders*, 27(3), 233–243. [PubMed: 9229256]
- Kolevzon A, Bush L, Wang AT, Halpern D, Frank Y, Grodberg D, ... Buxbaum JD (2014). A pilot controlled trial of insulin-like growth factor 1 in children with Phelan McDermid syndrome. *Molecular Autism*, 5, 54. [PubMed: 25685306]
- Lam KS, & Aman MG (2007). The Repetitive Behavior Scale-Revised: Independent validation in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 37(5), 855–866. [PubMed: 17048092]
- Leblond CS, Nava C, Polge A, Gauthier J, Huguet G, Lumbroso S, ... Bourgeron T (2014). Meta-analysis of SHANK mutations in autism spectrum disorders: A gradient of severity in cognitive impairments. *PLoS Genetics*, 10(9), e1004580. [PubMed: 25188300]
- Lord C, Rutter M, DiLavorne PC, Risi S, Gotham K, & Bishop SL (2012). Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1–4. Torrance, CA: Western Psychological Services.
- Luyster R, Qiu S, Lopez K, & Lord C (2007). Predicting outcomes of children referred for autism using the MacArthur–Bates Communicative Development Inventory. *Journal of Speech, Language, and Hearing Research*, 50(3), 667–681.
- Luyster RJ, Kadlec MB, Carter A, & Tager-Flusberg H (2008). Language assessment and development in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 38(8), 1426–1438. [PubMed: 18188685]
- Maloney ES, & Larrivee LS (2007). Limitations of age-equivalent scores in reporting the results of norm-referenced tests. *Contemporary Issues in Communication Science and Disorders*, 34, 86–93.
- McCathren RB, Yoder PJ, & Warren SF (1999). The relationship between prelinguistic vocalization and later expressive vocabulary in young children with developmental delay. *Journal of Speech Language and Hearing Research*, 42, 915–924.
- Mervis CB, & Robinson BF (2005). Designing measures for profiling and genotype/phenotype studies of individuals with genetic syndromes or developmental language disorders. *Applied Psycholinguistics*, 26(1), 41–64.
- Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, ... Scherer SW (2007). Contribution of SHANK3 mutations to autism spectrum disorder. *American Journal of Human Genetics*, 81(6), 1289–1297. [PubMed: 17999366]
- Mullen EM (1995). Mullen Scales of Early Learning. Circle Pines, MN: American Guidance Service.
- Phelan MC, Rogers RC, Saul RA, Stapleton GA, Sweet K, McDermid H, ... Kelly DP (2001). 22q13 deletion syndrome. *American Journal of Medical Genetics*, 101, 91–99. [PubMed: 11391650]
- Philippe A, Boddaert N, Vaivre-Douret L, Robel L, DanonBoileau L, Malan V, ... Munnich A (2008). Neurobehavioral profile and brain imaging study of the 22q13.3 deletion syndrome in childhood. *Pediatrics*, 122, e376–e382. [PubMed: 18625665]
- Plumb AM, & Wetherby AM (2013). Vocalization development in toddlers with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, 56, 721–734.

- Rapin I, & Dunn M (1997). Language disorders in children with autism. *Seminars in Pediatric Neurology*, 4(2), 86–92. [PubMed: 9195665]
- Rapin I, & Dunn M (2003). Update on the language disorders of individuals on the autistic spectrum. *Brain Dev*, 25(3), 166–172. [PubMed: 12689694]
- Rogers SJ, & Vismara LA (2008). Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child and Adolescent Psychology*, 37(1), 8–38. [PubMed: 18444052]
- Sakai Y, Shaw CA, Dawson BC, Dugas DV, Al-Mohtaseb Z, Hill DE, & Zoghbi HY (2011). Protein interactome reveals converging molecular pathways among autism disorders. *Science Translational Medicine*, 3(86), 86ra49.
- Schoen E, Paul R, & Chawarska K (2011). Phonology and Vocal Behavior in Toddlers with Autism Spectrum Disorders. *Autism Research*, 4(3), 177–188. [PubMed: 21308998]
- Soorya L, Kolevzon A, Zweifach J, Lim T, Dobry Y, Schwartz L, ... Buxbaum JD. (2013). Prospective investigation of autism and genotype phenotype correlations in 22q13 deletion syndrome and SHANK3 deficiency. *Molecular Autism*, 4, 18. [PubMed: 23758760]
- Sparrow SS, Cicchetti DV, Balla DA (2005). *Vineland-II Survey Forms Manual (Vineland Adaptive Behavior Scales)*. 2nd edition. Minneapolis, MN: AGS Publishing.
- Szatmari P, Bryson SE, Boyle MH, Streiner DL, & Duku E (2003). Predictors of outcome among high functioning children with autism and Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 44(4), 520–528. [PubMed: 12751844]
- Tager-Flusberg H, & Caronna E (2007). Language disorders: Autism and other pervasive developmental disorders. *Pediatric Clinics of North America*, 54(3), 469–481. [PubMed: 17543905]
- Tager-Flusberg H, Paul R, & Lord C (2005). Language and communication in autism. *Handbook of Autism and Pervasive Developmental Disorders*, 1, 335–364.
- Tager-Flusberg H, Rogers S, Cooper J, Landa R, Lord C, Paul R, ... Yoder P. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, 52(3), 643–652.
- Watt N, Wetherby A, & Shumway S (2006). Prelinguistic predictors of language outcome at 3 years of age. *Journal of Speech, Language, and Hearing Research*, 49, 1224–1237.
- Xu D, Yapanel U, Gray S (2009). Reliability of the LENA Language Environment Analysis System in young children's natural home environment. LENA Foundation Technical Report LTR05–02). Retrieved from <http://www.lenafoundation.org/TechReport.aspx/Reliability/LTR-05-2>.
- Xu D, Yapanel U, Gray S, Baer CT (2008). The LENA Language Environment Analysis System: The interpretive time segments (ITS) file. LENA Foundation Technical Report No. LTR-04–2). Retrieved from http://www.lenafoundation.org/TechReport.aspx/ITS_File/LTR-04-2.

Fig. 1. Using Transcriber software, the human transcriber demarcated sound segment boundaries for all speakers (Other Child (CXN), Near Adult Female (NAF)), other noises, and periods of silence (SIL). Each sound segment was coded as belonging to one of the eight possible sound source categories. Key Child sound segments were coded as speech-related vocalizations (CV) or non-speech sounds (CNS) including fixed signals and vegetative sounds

Fig. 2.

Pearson correlations were used to assess the relationship between LENA and human transcriber estimates of three potential expressive language measures. **a** Estimates of total Key Child segments were moderately correlated ($r = 0.707$, $p = 0.001$). **b** Estimates of child vocalizations were moderately correlated ($r = 0.60$, $p = 0.008$). **c** Estimates of the vocalization ratio were highly correlated ($r = 0.953$, $p < 0.001$)

Fig. 3.

Bland–Altman plots were constructed to visualize the agreement between LENA and human estimates of three potential expressive language measures. All data was log-transformed before conducting the analysis. The X-axis represents the average of the two estimates ($(\text{LENA} + \text{human transcriber})/2$). The Y-axis represents the difference or error between the two estimates ($\text{LENA} - \text{human transcriber}$). The *bold line* represents the estimated bias. *Dotted lines* represent the 95% limits of agreement (LoA). **a** The Bland–Altman plot of the Key Child counts reveals a mean difference of -0.296 , 95% LoA $[-0.667, 0.075]$. **b** The

Bland–Altman plot of the child vocalization counts reveals a mean difference of -0.248 , 95% LoA $[-0.751, 0.255]$. **c** The Bland–Altman plot of the vocalization ratios reveals minimal bias with a mean difference of -0.038 , 95% LoA $[-0.775, 0.700]$ count, however there was significant variability across participants. For one participant, LENA only identified 13% of the child vocalizations detected by the human transcriber. For another participant, however, the LENA child vocalization estimate was 257% of the human transcriber estimate. The Bland–Altman plot of the log-transformed data revealed a mean difference of -0.248 , 95% LoA $[-0.751, 0.255]$ (Fig. 3). There was a significant difference between the means of the two methods, $t(17) = -0.248$, $p = 0.001$.

Table 1

Participant characteristics: age, ASD symptomatology, expressive language scores, developmental quotients

Variable	N	<i>M</i> ± <i>SD</i>	Range
Age in months	18	76.92 ± 31.78	30–172
ADOS-2 total score	18	19.33 ± 4.102	10–25
Vineland expressive language raw score	18	15.44 ± 10.79	5–39
Mullen expressive language raw score	18	9.56 ± 5.66	4–25
Mullen nonverbal developmental quotient ^a	18	21.96 ± 11.06	9–45
Mullen verbal developmental quotient ^b	18	13.96 ± 6.45	5–28

^a Mullen developmental quotients (DQ) were calculated by dividing age equivalent scores on each domain by the child's chronological age and multiplying by 100. The nonverbal developmental quotient (NVDQ) includes the visual reception and fine motor subdomains

^b The verbal developmental quotient (VDQ) includes the receptive language and expressive language subdomains

Table 2

LENA-human inter-rater agreement: Key Child versus all other sound sources

Human transcriber	LENA system	
	Key Child	All other sound sources
Key Child	12,229 (40%)	18,127 (60%)
All other sound sources	2662 (3%)	86,942 (97%)

Table 3

LENA-human inter-rater agreement: child vocalizations vs non-speech sounds

Human transcriber	LENA system	
	Child vocalizations	Child non-speech sounds
Child vocalizations	6800 (84%)	1313 (16%)
Child non-speech sounds	1609 (39%)	2507 (61%)

Table 4Correlation of LENA accuracy measures and child-specific predictor variables^a

	Key Child interrater agreement	Child vocalization inter-rater agreement
Age		
Pearson correlation	−0.618 [*]	−0.577 [*]
Sig. (2-tailed)	0.006	0.012
ADOS-2, Module 1		
Pearson correlation	0.041	−0.204
Sig. (2-tailed)	0.871	0.417
ABC inappropriate speech		
Pearson correlation	−0.247	−0.567 [*]
Sig. (2-tailed)	0.323	0.014
RBS		
Pearson correlation	−0.140	−0.242
Sig. (2-tailed)	0.581	0.333

^{*}
 $p < 0.05$

^aThese findings are presented unadjusted for multiple comparisons

Table 5

Expressive language measures: original and logtransformed measurements

Variable	N	Transcriber <i>M(SD)</i>	LENA <i>M(SD)</i>	Mean difference <i>M(SD)</i>
Original measurements				
Key Child count	18	1686.44 (1107.17)	828.11 (406.57)	858.33 (868.84)
Child vocalization count	18	1060.78 (625.67)	563.72 (261.54)	497.06 (513.19)
Vocalization ratio	18	4.80 (6.16)	2.91 (2.41)	1.88 (3.93)
Log-transformed measurements				
Key Child count	18	3.17 (0.22)	2.87 (0.22)	0.29 (0.19)
Child vocalization count	18	2.95 (0.27)	2.70 (0.22)	0.25 (0.26)
Vocalization ratio	18	0.399 (0.582)	0.361 (0.296)	0.04 (0.38)