Title

Predicting Expressive Language From Early Vocalizations in Young Children With Autism Spectrum Disorder: Which Vocal Measure Is Best?

Permalink

Journal

ISSN

Authors

McDaniel, Jena
Yoder, Paul
Estes, Annette
et al.

Publication Date

DOI

Peer reviewed

Predicting Expressive Language From Early Vocalizations in Young Children With Autism Spectrum Disorder: Which Vocal Measure Is Best?

Jena McDaniel,[a] [iD] Paul Yoder,[b] Annette Estes,[c] and Sally J. Rogers[d]

Purpose: This study was designed to test the incremental validity of more expensive vocal development variables relative to less expensive variables for predicting later expressive language in children with autism spectrum disorder (ASD). We devote particular attention to the added value of coding the quality of vocalizations over the quantity of vocalizations because coding quality adds expense to the coding process. We are also interested in the added value of more costly human-coded vocal variables relative to those generated through automated analyses.
Method: Eighty-seven children with ASD aged 13–30 months at study initiation participated. For quantity of vocalizations, we derived one variable from human coding of brief communication samples and one from an automated process for daylong naturalistic audio samples. For quality of vocalizations, we derived four human-coded variables and one automated variable. A composite expressive language measure was derived at study entry, and 6 and 12 months later. The 12 months–centered intercept of a simple linear growth trajectory was used to quantify later expressive language.
Results: When statistically controlling for human-coded or automated quantity of vocalization variables, human-coded quality of vocalization variables exhibited incremental validity for predicting later expressive language skills. Human-coded vocal variables also predicted later expressive language skills when controlling for the analogous automated vocal variables.
Conclusion: In sum, these findings support devoting resources to human coding of the quality of vocalizations from communication samples to predict later expressive language skills in young children with ASD despite the greater costs of deriving these variables.
Supplemental Material: https://doi.org/10.23641/asha.12276458

Although most individuals with autism spectrum disorder (ASD) exhibit difficulties with language, the nature and severity of these difficulties vary widely (e.g., Kjelgaard & Tager-Flusberg, 2001; Lord et al., 2004; Tager-Flusberg & Joseph, 2003; Tager-Flusberg et al., 2005; Thurm et al., 2007). Approximately 30% of individuals with ASD use few or no spoken words despite years of spoken language intervention (Anderson et al., 2007;

Tager-Flusberg & Kasari, 2013). In contrast, other individuals with ASD are verbally fluent with large vocabularies and complex syntax (Kjelgaard & Tager-Flusberg, 2001; Tager-Flusberg & Joseph, 2003). For children with ASD who are not yet talking, strong early predictors of language development would be useful for planning intervention. Measures of vocalizations, which precede spoken words, are likely candidates for such predictors. Although replicated findings show a strong correlation between the quantity and/or quality of vocalizations (i.e., voiced sounds produced by the vocal folds on exhalation) and expressive language in children with ASD (see McDaniel et al., 2018, for a meta-analytic review), which vocal variables are strongest predictors is unknown. Identifying vocal variables that are especially predictive of later expressive language in initially preverbal or low verbal children with ASD may help identify how or for whom language intervention works

[a]Life Span Institute, University of Kansas, Lawrence
[b]Department of Special Education, Vanderbilt University, Nashville, TN
[c]Department of Speech and Hearing Sciences, University of Washington, Seattle
[d]Department of Psychiatry and Behavioral Sciences, University of California, Davis

Correspondence to Jena McDaniel: jena.c.mcdaniel@ku.edu

or explain associations between caregiver responses to child vocalizations and expressive language. Optimal vocal measures balance cost and utility in a parsimonious manner.

Differences in Cost Based on Vocalization Variable Type

Vocal variables can be very simple or quite complex. Quantity measures are relatively simple. They count the number of vocalizations regardless of the vocalizations' contents or use. Measures of the quality of vocalizations are relatively more complex because they consider features that influence message saliency and clarity. These features include the communicative quality (e.g., higher communicative quality shown by directing a vocalization toward another person with eye contact than by not directing such vocalization) and phonological quality of vocalizations (e.g., higher phonological quality shown by including consonants or canonical syllables in a vocalization than by omitting them). Quality variables require more time to train observers and for coders to complete the coding tasks. Thus, quality variables are more expensive than quantity variables.

Vocal variables can be derived from human coding of brief communication samples or automated (i.e., computer-generated) analyses of daylong vocal samples. Human-coded vocal variables require substantial time to train observers, code video-recorded communication samples, check for observer drift, and estimate interobserver reliability. Automated analyses of daylong vocal samples are less costly than human coding in terms of personnel time. The automated vocal analysis process includes collecting and processing the audio recordings using specialized technology (see Daylong Naturalistic Audio Samples for Automated Vocal Variables section for details), which requires limited training to operate. No person needs to listen to any of the audio samples, and an entire participant sample's data can be processed in only a few hours. In contrast, it can take a few hours to hand code vocal variables from only a couple of communication samples. Automated analysis costs are primarily financial (e.g., equipment, recording devices, specialized clothing to hold recording devices, and software licenses). It should be noted that the communicative quality of vocalizations cannot be drawn from automated means because the available technology cannot infer communicative intent or directedness of vocalizations. Even though automated analyses cannot provide information on communicative quality, they can provide information on the quantity and phonological quality of children's vocalizations.

Considering Variable Cost and Utility to Maximize Resources

Given the variation in vocal variable types and analysis methods, investigators can maximize resources for analyzing vocalizations by using variables that are easy to code (e.g., quantity) or automated variables from daylong, naturalistic vocal samples. However, if the quality of vocalizations predicts expressive language better than quantity, then using extra resources to derive measures of quality of vocalizations might be justified. Similarly, if human coding increases the extent to which we can predict expressive language compared with automated vocal variables, then human-coded communication samples might be justified. One way to assess added value is to examine whether a more costly vocal variable predicts later expressive language after controlling for a less costly vocal variable (i.e., incremental validity). We take this approach in this study using variables with theoretical and empirical support.

Theoretical and Empirical Support for Vocal Development Predicting Expressive Language

Current evidence supports the continuity of prelinguistic vocalizations (e.g., babbling) and spoken words. This supportive evidence includes individual children producing the same phonemes in prelinguistic vocalizations as early words (McCune & Vihman, 2001; Oller, 2000; Vihman, 2017; Vihman et al., 1985), language-specific acoustic characteristics of vocalizations (Oller, 2000; Rvachew et al., 2006), and the quantity and quality of vocalizations (e.g., inclusion of consonants and canonical syllables) predicting expressive language in children with typical development (e.g., Stoel-Gammon, 1991; Watt et al., 2006).

The continuity between vocalizations and spoken words aligns with at least three theories of language development that emphasize bidirectional child–caregiver interactions in facilitating vocal and language development, including the social feedback theory (Goldstein et al., 2003; Goldstein & Schwade, 2008), social feedback loop theory (Warlaumont et al., 2014), and transactional theory of spoken language development (Camarata & Yoder, 2002; McLean & Snyder-McLean, 1978; Sameroff & Chandler, 1975; Woynaroski et al., 2014). Child-driven theories of spoken language development also support the continuity between vocalizations and expressive language based on shared articulators (e.g., tongue and lips) and motor movements for producing prelinguistic vocalizations and words (Fry, 1966; Iverson, 2010; Stoel-Gammon, 2011; Vihman, 1992, 1996). For example, producing "Bah" in a vocalization without lexical meaning and producing "Bah" as an approximation of "ball" use the same motor movements of the articulators.

Empirically, concurrent and longitudinal correlations have been reported between the quantity and quality of vocalizations and expressive language in children with ASD. The quantity of vocalizations has correlated with expressive language impairment concurrently and predictively in children with ASD for human-coded variables (Plumb & Wetherby, 2013) and automated analyses (Dykstra et al., 2013). For example, the quantity of total vocalizations correlated at the same time with the Communication and Symbolic Behavior Scales (CSBS) Speech Composite score, a measure of spoken expressive language and speechlike vocalizations ($r = .47$; Plumb & Wetherby, 2013) for 18- to 24-month-old children with ASD. The correlation between the quantity of vocalizations during the second year of

life and verbal developmental quotient at age 3 years was also significant for children with ASD ($r = .39$; Plumb & Wetherby, 2013). Using automated analyses, the quantity of child speech–related vocalizations correlated ($r = .33$) with language skills for 3- to 5-year-old children with ASD (Dykstra et al., 2013). In contrast, the quantity of vocalizations per hour did not correlate significantly with expressive language concurrently for children with ASD with a mean chronological age of 76.92 months ($SD = 31.78$ months; Rankine, 2016).

Communicative quality of vocalizations has predicted expressive language in children with ASD (Plumb & Wetherby, 2013) as well, but again, not universally (Swineford, 2011). Plumb and Wetherby (2013) reported that communicative vocalizations in the second year of life predicted expressive language skills at age 3 years above and beyond noncommunicative vocalizations. In contrast, Swineford (2011) reported nonsignificant correlations between communication acts with vocalizations within home observations and the CSBS Words subscale ($r = .03$) or the CSBS Speech Composite ($r = .13$) concurrently for children suspected of having ASD (mean chronological age = 19.51 months, $SD = 2.34$ months).

Phonological quality has correlated with current and later expressive language skills of children with ASD when using human-coded (Book, 2009; McCoy, 2013; McDaniel et al., 2019; Talbott, 2014; Wetherby et al., 2007; Woynaroski et al., 2017; Yoder et al., 2015) and automated vocal variables (e.g., use of consonants or canonical syllables; Woynaroski et al., 2017). For human-coded phonological quality, the rate of canonical babbling correlated with concurrent expressive language ($r = .65$) in children with ASD (mean chronological age = 44.67 months, $SD = 8.35$ months; Sheinkopf et al., 2000). Additionally, Yoder et al. (2015) found that the inventory of consonants used in communication acts predicted expressive language growth in initially preverbal children with ASD over and above 10 other putative predictors. Similarly, Wetherby et al. (2007) identified that inventory of consonants used in communication acts at ages 18–24 months was one of the "best predictors of verbal skills at 3 years" (p. 971), compared with numerous other possible predictors for children with ASD. For automated phonological quality variables, the average count per utterance–consonants + vowels (ACPU-C+V) predicted expressive language 4 months later ($r = .55$; Woynaroski et al., 2017).

It is unclear whether the associations between vocal quantity and expressive language are due to the intercorrelation of vocal quantity and quality in children with ASD. One way to address these issues is to examine the incremental validity of each in predicting expressive language in children with ASD.

Research Questions

Prior investigations have not compared the relative predictive utility of more expensive vocal variables (e.g., human-coded quality variables) to variables that are less costly (e.g., automated and quantity variables). We examine two research questions of the incremental validity of vocal variables in young children with ASD.

1. After statistically controlling for quantity of vocalizations, does quality of vocalizations account for unique variance in later expressive language skills?

2. After statistically controlling for an automated measure of the same vocalization aspect (i.e., quantity or phonological quality), do human-coded variables account for unique variance in later expressive language skills?

Method

Caregivers provided written informed consent prior to participants beginning the study. Institutional review boards at the University of California at Davis, the University of Washington, and Vanderbilt University approved all study procedures.

Participants

Eighty-seven children (21 girls, 66 boys) from a multisite randomized controlled trial participated in this study (Rogers et al., 2013). Although the participants were randomly assigned to two treatment styles and treatment intensities, the results of the current analyses were not influenced by group membership. That is, Predictor × Group interactions were nonsignificant. Therefore, treatment styles and intensities are not discussed further.

For inclusion, participants had to be 13–30 months of age at study entry, meet the criteria for ASD on multiple measures (i.e., *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* [American Psychiatric Association, 2013], Autism Diagnostic Interview–Revised [Lord et al., 1994], and Autism Diagnostic Observation Schedule for Toddlers [Luyster et al., 2009]), achieve an overall developmental quotient of at least 35 on the Mullen Scales of Early Learning (MSEL; mental age/chronological age × 100; Mullen, 1995), live in a home that uses spoken English at least 60% of the time per caregiver report, walk without primary motor impairments affecting hand use, and have hearing and visual acuity within normal limits. Participants were not excluded based on the presence of genetic disorders or other health conditions. Participants had a mean chronological age of 23.42 months at study entry ($SD = 3.98$) and a mean developmental quotient of 58.83 ($SD = 17.96$). On the MSEL, participants presented with a mean age equivalent of 10.11 months ($SD = 7.22$) on the Receptive Language subscale and 11.97 months ($SD = 4.71$) on the Expressive Language subscale.

Forty-eight participants were reported to be White, 19 to be more than one race, nine to be Asian, seven to be Black or African American, one to be American Indian or Alaskan native, one to be Native Hawaiian or other Pacific Islander, and two as unknown. Seventeen participants

were reported to be Hispanic/Latino, 64 to be non-Hispanic, and six as unknown. One mother had some high school education, six had a high school diploma, 25 had some college education, 24 had a college degree, six had some graduate school education, 22 had a graduate degree, and one reported "other."

Procedure

Table 1 displays the study's constructs, procedures, and variables. Data are used from procedures administered across three time periods that spanned 12 months (Time 1 = study initiation; Time 2 = 6 months poststudy initiation; Time 3 = 12 months poststudy initiation).

Communication Samples for Human Coding of Vocal Variables

Each human-coded vocal variable was derived from a 15-min Communication Sample Procedure (CSP) and three 6-min Early Communication Indicator (ECI) sessions (Greenwood et al., 2006; Luze et al., 2001) at Time 1. Because the ECI is brief, we averaged scores from the first 3 months to increase stability of variable scores (Yoder et al., 2018). Averaging across the sessions permitted inclusion of participants who were missing one (n = 14) or two ECI samples (n = 3) for Months 1–3. The use of three ECI samples per time point provided a total of 18 min per time point per participant, similar to the 15-min CSP sample, and increased the stability of the vocal variables. Both communication sampling contexts are play-based interactions in which the child engages with an examiner using a standard toy set. The examiner uses responsive interaction style principles when engaging with the child to support engagement (e.g., follow the child's lead and join in and play at the child's demonstrated level of play) and communication (e.g., talking about topics related to the child's focus of attention, monitoring examiner utterance length and complexity, and avoiding directives).

Daylong Naturalistic Audio Samples for Automated Vocal Variables

Participants' families collected one daylong audio recording at study initiation with the LENA digital recording device (LENA Research Foundation, 2015). The LENA digital recording device was placed in a specialized vest's pocket for the participant to wear for 12–16 hr of recording. No specifications were given regarding the day of the week or setting for the recording, except to avoid the participant being ill or going swimming. Caregivers were instructed to remove the participant's vest, with the recorder still on, and place it near the participant when he or she was sleeping or in the car. Trained research assistants downloaded the digital audio files from the returned recording devices to a designated computer for processing and analysis.

Expressive Language Procedures

Variables derived from four procedures were used for the expressive language composite variable: number of different root words said from the CSP, Expressive subscale age-equivalency score on the MSEL, raw score for words said on the MacArthur–Bates Communicative Development Inventory compilation form (MB-CDI; Fenson et al., 2006), and Expressive Language subscale age-equivalency score on the Vineland Adaptive Behavior Scales–Second Edition (VABS; Sparrow et al., 2005). Thus, two caregiver reports (i.e., MB-CDI and VABS) and two direct observations (i.e., CSP and MSEL) were used. To calculate the number of different root words said from the CSP, trained coders transcribed all intelligible words produced by the participants in each communication sample using ProcoderDV software (Tapp, 2003). Then, the coder used Systematic Analysis of Language Transcripts software (Miller & Chapman, 2016) to calculate the number of different root words said. The coding manual is available from the first author. On the MB-CDI, the caregiver indicates which words from a combined list of the Words and Gestures and the Words and Sentences forms (total words = 720) that the child has said in the prior 2 weeks. The MB-CDI, VABS, and MSEL were administered at all three time points. The CSP was only administered at Times 1 and 3 due to design and budget constraints.

Observational Coding of Communication Samples

Trained research assistants and the first author completed observational coding for the CSP and ECI using ProcoderDV and Systematic Analysis of Language Transcripts

Table 1. Matrix of vocal variables by construct and derivation method.

Construct	Human-coded variables from communication samples	Automated variables from daylong naturalistic audio samples
Quantity	Number of total vocalizations	AQV
Communicative quality	Number of communication acts with a vocalization	Not applicable
	Proportion of communicative vocalizations	
Phonological quality	DKCC	ACPU-C+V
	Proportion of vocalizations with a canonical syllable	

Note. Variables within the same cell are combined to form composites for each construct. AQV = automated quantity of vocalizations (which is the number of child vocalizations); DKCC = diversity of key consonants used in communication acts (Wetherby et al., 2007; Woynaroski et al., 2017); ACPU-C+V = average count per utterance–consonants + vowels (Xu et al., 2014).

software. Coders completed four passes using timed event behavior sampling to code behaviors used to derive the vocal variables. The coder first identified codable and uncodable (i.e., the child's face is not visible for at least 10 s) portions of each video file. The coder identified all communication acts within the codable time on the second pass. The coding manual, which is available from the first author, includes detailed coding rules for codable portions of the communication sample and for communication acts. Conceptually, communication acts were defined as a spoken/signed word or a nonword vocalization or gesture with coordinated attention to object and communication partner.

On the third pass, the coder identified and classified vocalizations within communication acts. Vocalizations were defined as nonvegetative voiced sounds (i.e., created by vibrating vocal folds) created during exhalation. Voiced laughs, voiced sighs, voiced cries, whispered productions, isolated voiceless consonants, glottal fry, ingressive phonation, and reflexive, vegetative sounds (e.g., sounds from burps, hiccups, coughs, sneezes, throat clearing, tongue clicking, and lip popping) were not coded as vocalizations. During this pass, the coder also indicated whether each vocalization contained one or more codable consonants (i.e., /m/, /n/, /b/ or /p/, /d/ or /t/, /g/ or /k/, /w/, /l/, "y," /s/, and "sh") and/or a canonical syllable. Canonical syllables had to include at least one consonant, at least one vowel, and a quick, uninterrupted transition from the consonant to vowel or vowel to consonant.

Because coding for the earlier passes focused on communication acts, a fourth pass was necessary to code presence of noncommunicative vocalizations and whether these included codable consonants or canonical syllables. CSP and ECI session variables from the same time period were averaged after checking for a sufficiently high correlation between them.

Vocal Variables

Human-Coded Vocal Variables
From Communication Samples

Table 1 displays all of the vocal variables. For quantity, we calculated the number of total vocalizations produced by the child during the communication sampling procedures. For communicative quality, we calculated (a) the number of communication acts with a vocalization and (b) proportion of communicative vocalizations (i.e., number of communicative vocalizations divided by the number of total vocalizations). For phonological quality, we coded (a) the diversity of key consonants used in communication acts (DKCC) and (b) the proportion of vocalizations with a canonical syllable (i.e., number of vocalizations with a canonical syllable divided by the number of total vocalizations). DKCC is the number of 10 specific consonants (i.e., /m/, /n/, /b/ or /p/, /d/ or /t/, /g/ or /k/, /w/, /l/, "j," /s/, and "sh") used communicatively (Wetherby et al., 2007; Woynaroski et al., 2017). Because members of voiced–voiceless pairs (e.g., /b/ vs. /p/, /d/ vs. /t/, /g/ vs. /k/) are

difficult to distinguish reliably on recordings, children can only receive 1 point for the pair regardless of whether they produce one or both of each pair's consonants.

Automated Vocal Variables From Daylong Audio Samples

For quantity of vocalizations, we used the automated quantity of vocalizations (AQV). AQV is the total number of child vocalizations from the entire audio sample, which is available in the standard LENA Pro software package. Speech segments produced by the child wearing the recorder that are preceded and followed by a pause of at least 300 ms are counted as child vocalizations. Nonspeech sounds (e.g., vegetative sounds, cries, and other fixed-signal sounds) are not included. See Oller et al. (2010) and Xu et al. (2008) for a detailed description of how the LENA system segments acoustic events and determines the sound source for each. Regarding reliability of automated analyses, a number of investigations have compared the degree to which automated analyses identify and classify vocalizations similar to human coders (e.g., Rankine, 2016; VanDam & Silbert, 2016; Xu et al., 2014). For example, automated analyses accurately identified more than 72% of clear child utterances in a sample of children aged 2–48 months (Xu et al., 2014). VanDam and Silbert (2016) also reported a high agreement between automated analyses and human coding of vocalizations of children with typical development (M_{age} = 29.1 months).

For phonological quality of vocalizations, we used ACPU-C+V scores. These scores are based on Sphinx speech recognition software, which is modeled on adult data. Sphinx software estimates how many times certain types of phones occur within speech-related utterances, which are further categorized into consonants versus vowels, and silence within utterances and other nonspeech sounds (e.g., hesitation, coughing, noise, and lip smacking; Xu et al., 2014). The phones identified through the Sphinx speech recognition software are broader than English phonemes. Average count per utterance–consonants (ACPU-C) and average count per utterance–vowels (ACPU-V) have been found to correlate highly with human coding, $r = .85$ and $r = .82$, respectively; however, the automated analyses were more conservative than the human coding and underestimated the counts of consonants and vowels (Xu et al., 2014). For more information about the validation process and comparisons with human coding, see Rankine (2016), VanDam and Silbert (2016), and Xu et al. (2008, 2014). ACPU-C and ACPU-V scores had to be derived using computer programs housed at the LENA Research Foundation because the necessary software is not commercially available. The system does not reliably identify specific consonant or specific vowel sounds in young children with ASD; thus, consonant or vowel tokens, not types, are reflected in the counts. We then created a z score composite of these ACPU-C and ACPU-V scores to create ACPU-C+V as a more stable estimate of the phonological quality of vocalizations.

Results

Preliminary Analyses

Descriptive Statistics

Supplemental Material S1 displays the means, standard deviations, and ranges for the vocal variables. Results for human-coded vocal variables are presented by procedure. Correlations between vocal variables are shown in Supplemental Material S2.

Interobserver Reliability

For each time point for variables derived from the CSP and ECI, a trained secondary coder independently coded a random sample of at least 20% of coded sessions. The primary coder was blind to which sessions would be coded for reliability. The study's analyses use the primary coder's coding. Two-way mixed effects single measures intraclass correlation coefficients (ICCs) with absolute agreement account for differences in unitizing and classifying behaviors between coders and for the variance among participants on the component variables addressing the research questions.

Reliability values for each procedure and time period all exceeded our benchmark of .70 for "very good" (Mitchell, 1979). For interobserver reliability, the human-coded vocal variables had a mean ICC of .92 (SD = .07) across both communication sampling procedures (i.e., CSP and ECI). The vocal variables from the ECI showed sufficient stability across the three Time 1 sessions (mean ICC = .79, SD = .05). The number of different root words said in the CSP (one component of the expressive language composite) had an ICC of .77 for Time 1 across coders.

Creating Composite Variables

To increase stability and reduce the number of variables for the incremental validity analyses, we created composites by averaging the z score–converted raw scores. We used a threshold of $r \geq .40$ for sufficient correlations between component variables for these composite variables (Cohen & Cohen, 1984; Yoder et al., 2018).

For the expressive language composite, we used the sample's Time 3 (12 months after study initiation) mean and standard deviation to permit growth across time periods. All of the expressive language component variables correlated sufficiently (rs = .52–.77 at Time 1, rs = .64–.87 at Time 2, and rs = .69–.85 at Time 3). Thus, the number of different root words said from the CSP, MSEL Expressive subscale age-equivalency scores, MB-CDI compilation form raw scores for words said, and VABS Expressive Language subscale age-equivalency scores were converted to z scores individually and then averaged to create the expressive language composite.

For the human-coded vocal variables, we created composites from component variables coded from the CSP and ECI. Correlations between values from the CSP and ECI were sufficiently correlated for all variables (rs = .60–.80). Thus, the z scores for one CSP sample and three ECI samples were averaged for the human-coded vocal variables. We also created composites of human-coded

component variables assessing the same quality feature (i.e., communicative quality composite and phonological quality composite). Component variables were sufficiently correlated for communicative quality (r = .79) and phonological quality (r = .70).

For the automated vocal variables, ACPU-C and ACPU-V correlated strongly (r = .82). Thus, we computed average z-converted scores for ACPU-C and ACPU-V to represent the automated measure of the phonological quality of vocalizations (i.e., ACPU-C+V) as planned.

Modeling Growth of Expressive Language

We sought to test whether Time 1 vocal variables predicted variation in this best available estimate of expressive language at Time 3. In growth curve models in which time is centered at the end point, the intercept is typically a better estimate of the end point expressive language than the observed component or composite variables because the intercept is based on information from multiple data points rather than a single time point (Singer & Willett, 2003). Because time was centered at Time 3 in the current study, the intercept of the growth model for the expressive language composite represented the best available estimate of expressive language at Time 3. Missing data varied from 1% to 16%, depending on the variable and measurement period, M = 9%, SD = 5%. We used full maximum likelihood estimation to address missing data (Enders, 2010). For model selection, we used a buildup approach that progressed from the simplest (but possibly ill fitting) model to increasingly more complex (but better fitting) models, accepting the more complex model only if a chi-squared test indicated improved model fit assuming the parameters are nonredundant (r between parameters < .90). The random intercept, fixed slope model provided evidence of a better fit than a fixed intercept, fixed slope model, as evidenced by a lower −2 log likelihood value (i.e., 633 vs. 421). The correlation between the intercept and slope in the latter model was .45. Although the −2 log likelihood value decreased further for the random intercept, random slope model relative to the random intercept, fixed slope model, the correlation between slope and intercept was very high (r = .92). The high covariance between the intercept and slope means that there is limited variance in slope remaining to be explained after controlling for intercept. Due to this high covariance of the intercept and slope and the desire to use a well-fitted yet parsimonious growth model, we chose to use the random intercept, fixed slope model. Thus, the intercept from the random intercept, fixed slope model was used as the best estimate of end point (Time 3) expressive language. For these and other growth curve models, the data met the statistical assumptions of homoscedasticity and residuals fell within the acceptable parameters for skewness (< |.8|) and kurtosis (< |3.0|; Tabachnick & Fidell, 2001).

Zero-Order Associations With Later Expressive Language

In separate models, we added each vocal variable (not composite variables) to a model predicting the end

point–centered intercept of the expressive language composite. All vocal variables significantly and positively predicted expressive language. As shown in Table 2, all of the coefficients are significant. Notably, all of the human-coded variables except the quantity of vocalizations variable (i.e., number of total vocalizations) exhibited a large effect size (i.e., pseudo $R^2 \geq .25$) for predicting expressive language skills 12 months later. In contrast, neither of the automated measures exhibited a large effect size. The pseudo R^2 value provides an effect size that represents the amount of explainable variance accounted for by the predictor variable (Xu, 2003).

Incremental Validity: Quantity Versus Quality

We examined whether the vocal quality composite variables explained additional variance in expressive language after controlling for each of the vocal quantity component variables. The coefficients for quantity and quality are shown in Table 3 when testing the incremental validity of communicative quality of vocalizations and in Table 4 when testing the phonological quality of vocalizations. Results for each model are displayed in more detail in Supplemental Material S3. Each line in the table represents a separate model. When the coefficient for the predictor variable is significant, it accounts for unique variance in later expressive language after statistically controlling the other vocal variable in the model.

Human-Coded Communicative Quality Composite Controlling for Vocal Quantity

The communicative quality composite from the human-coded communication samples (i.e., number of communication acts with a vocalization and the proportion of communicative vocalizations) strongly predicted later expressive language regardless of which vocal quantity variable was controlled (see Table 3 and Supplemental Material S3). The effect size of the association of communicative quality composite predicting later expressive language was statistically significant and quite large when controlling for human-coded (pseudo R^2 change = .61) or automated (pseudo R^2 change = .62) vocal quantity, respectively.

Human-Coded Phonological Quality Composite Controlling for Vocal Quantity

The phonological quality composite from human-coded communication samples (i.e., DKCC and the proportion of vocalizations with a canonical syllable) was very strongly associated with later expressive language even after controlling for the quantity of vocalizations, regardless of how vocal quantity was measured (see Table 4 and Supplemental Material S3). The effect size of the association between human-coded phonological quality and later expressive language was .50 and .49, when controlling for human-coded or automatic vocal quantity, respectively.

Automated Phonological Quality Controlling for Vocal Quantity

ACPU-C+V predicted later expressive language after controlling for human-coded vocal quantity with an effect size of .10. In contrast, ACPU-C+V did not exhibit incremental validity after controlling for AQV with an effect size of .05. See Table 4 and Supplemental Material S3 for details.

Incremental Validity: Human-Coded Versus Automated

To evaluate the incremental validity of human-coded measures after controlling for the analogous automated measures, we added the measure from each of the two data collection and variable derivation methods that purportedly measured the same construct in the same model to predict end point expressive language. Table 5 provides the details of these analyses.

When the number of total vocalizations (human-coded) and AQV were in the same model, both were significant unique predictors of end point expressive language. Additionally, the human-coded quantity variable accounted for a medium amount of explainable variance in expressive language after controlling for AQV (pseudo R^2 change = .10).

When the automated phonological quality variable (ACPU-C+V) and the human-coded phonological quality composite were in the same model predicting later expressive language, only the human-coded phonological quality

Table 2. Zero-order fixed effects estimates for vocal variables predicting end point expressive language.

Construct	Vocal variable	Coeff.	SE	t	df	p	Pseudo R^2
Quantity	Number of total vocalizations[HC]	0.26	0.07	3.74	89	< .001	.16
	AQV[A]	2.9×10^4	6.8×10^5	4.35	84	< .001	.19
Communicative quality	Number of communication acts with a vocalization[HC]	0.86	0.10	8.64	86	< .001	.54
	Proportion of communicative vocalizations[HC]	0.69	0.07	9.35	86	< .001	.59
Phonological quality	DKCC[HC]	0.59	0.06	9.41	86	< .001	.60
	Proportion of vocalizations with a canonical syllable[HC]	0.38	0.05	7.26	86	< .001	.44
	ACPU-C+V[A]	0.27	0.10	2.71	87	.01	.08

Note. See Xu (2003) for pseudo R^2 details. Coeff. = unstandardized coefficient; *SE* = standard error; [HC] = human-coded; AQV = automated quantity of vocalizations (which is the number of child vocalizations); [A] = automated; DKCC = diversity of key consonants used in communication acts (Wetherby et al., 2007; Woynaroski et al., 2017); ACPU-C+V = average count per utterance–consonants + vowels (Woynaroski et al., 2017; Xu et al., 2014).

Table 3. Unstandardized coefficients (standard errors), significance, and effect size of the human-coded communicative quality composite variable predicting end point expressive language after controlling each vocal quantity component variable.

Number of total vocalizations[HC]	AQV[A]	Communicative quality composite[HC]	Pseudo R^2 change
−0.01 (0.06)		0.91 (0.10)***	.61
	9.7×10^5 (5.2×10^5)	0.83 (0.09)***	.62

Note. See Xu (2003) for pseudo R^2 details. [HC] = human-coded; AQV = automated quantity of vocalizations (which is the number of child vocalizations); [A] = automated.

***$p < .001$.

composite was a significant predictor. Adding the human-coded phonological quality composite to the model accounted for a large amount of the explainable variance in expressive language after controlling for the automated phonological quantity composite (pseudo R^2 change = .54). The incremental validity of the ACPU-C+V was nonsignificant after controlling for the human-coded phonological quality composite.

Discussion

When predicting later expressive language from vocal variables, incremental validity is arguably among the most rigorous methods to demonstrate whether it is worth the expense to quantify the quality of vocalizations rather than just the quantity and/or to train observers to code communication samples. In this study, we tested incremental validity for assessing the quality of vocalizations (communicative and phonological quality), which is relatively more expensive than coding the quantity of vocalizations. We also tested incremental validity of the more costly human-coded quantity and quality variables relative to less costly analogous automated variables for the quantity and phonological quality of vocalizations. Automated analyses require a large initial financial investment, but the personnel and time costs in using them are lower than human coding of communication samples. Therefore, over time, automated analyses are less expensive than human coding of vocalizations.

The findings support the use of human-coded variables of the quality of vocalizations for young children with ASD in the early stages of word learning, despite the increased costs, relative to less costly quantity variables or automated variables. Human-coded quality variables account for unique variance, with large effect sizes, in predicting expressive language in young children with ASD in the early stages of word learning, even after controlling for the information that the less expensive vocal variables provided.

Findings are consistent with prior findings regarding the predictive value of vocal variables. Prior studies have also found that expressive language is correlated with or predicted by the number of vocal communication acts (Plumb, 2008), DKCC (Wetherby et al., 2007; Woynaroski, 2014; Woynaroski et al., 2017; Yoder et al., 2015), and ACPU-C+V (Woynaroski et al., 2017), as well as incremental validity for DKCC (Yoder et al., 2015). However, Plumb and Wetherby (2013) reported a nonsignificant association between the proportion of communicative vocalizations and expressive language. This discrepancy with the current findings may be at least partially explained by the larger sample size, use of growth curve modeling with an expressive language composite (i.e., not a single measure at one point in time), and use of a composite communicative quality predictor in the current study.

Limitations

At least four limitations should be acknowledged. First, because validation refers to a specific variable, use, and population, replication is necessary to apply the findings to other variables, uses, or populations (Yoder et al., 2018). Second, the use of multiple *t* tests without alpha adjustment for a given research question increases risk for Type I errors. However, because a number of the findings (e.g., human-coded number of total vocalizations, DKCC,

Table 4. Unstandardized coefficients (standard errors), significance, and effect size of each phonological quality composite variable predicting end point expressive language after controlling for each vocal quantity component variable.

Number of total vocalizations[HC]	AQV[A]	Phonological quality composite[HC]	ACPU-C+V[A]	Pseudo R^2 change
−0.11 (0.07)		0.60 (0.08)***		.50
	9.5×10^5 (6.9×10^5)	0.47 (0.06)***		.49
0.27 (0.07)***			0.24 (0.09)**	.10
	2.6×10^4 (6.9×10^5)***		0.17 (0.09)	.05

Note. See Xu (2003) for pseudo R^2 details.

[A] = automated; ACPU-C+V = average count per utterance–consonants + vowels (Woynaroski et al., 2017; Xu et al., 2014). [HC] = human-coded; AQV = automated quantity of vocalizations;

**$p < .01$. ***$p < .001$.

Table 5. Coefficient (standard errors), significance, and effect size of the human-coded vocal variable predicting end point expressive language after controlling for the automated vocal variable.

Automated vocal variable		Human-coded vocal variable		
Variable	Coefficient	Variable	Coefficient	Pseudo R^2 change
AQV	2.2×10^4 (7.1×10^5)**	Number of total vocalizations	0.19 (0.07)**	.10
ACPU-C+V	0.07 (0.08)	Phonological quality composite	0.50 (0.06)***	.54

Note. See Xu (2003) for pseudo R^2 details. AQV = automated quantity of vocalizations; ACPU-C+V = average count per utterance–consonants + vowels (Woynaroski et al., 2017; Xu et al., 2014).

**$p < .01$. ***$p < .001$.

and ACPU-C+V predicting expressive language; Plumb, 2008; Woynaroski, 2014; Woynaroski et al., 2017; Yoder et al., 2015) are replications, the risk of findings being solely due to Type I errors decreases. Novel findings (e.g., AQV, the proportion of communicative vocalizations, and the proportion of vocalizations with a canonical syllable predicting expressive language, as well as the incremental validity findings) require replication. Third, relative cost as a general consideration was a motivation for the current study. However, limited resources prevented detailed cost analyses. Finally, because this study used a correlational design, we cannot eliminate all third variable explanations for the association between vocal variables and later expressive language.

Strengths

Four strengths should be acknowledged. First, this study includes human-coded and automated vocal variables for the same participants, which enables direct comparisons of the predictive validity of two ways to measure vocal quantity and quality. Second, the use of multilevel modeling provides a better estimate of end point expressive language than relying on the observed value (Singer & Willett, 2003). Third, the relatively large sample size ($N = 87$) permitted sufficient power for detecting incremental validity. Fourth, the relatively long study duration (i.e., 12 months) enables prediction of expressive language from early vocal variables across this relatively long period of time. Twelve months is a meaningful interval for this type of predictive study because intervention goals are often written for yearlong intervals.

Future Directions

Limited resources prevented detailed cost analyses for the current study. Future studies should consider the monetary cost of variables to inform variable selection and planning of later investigations. Relatedly, the reliability of live coding the communicative and phonological quality of vocalizations and the amount of training time required to achieve adequate reliability warrant investigation. Such information informs which variables may be most feasible in clinical settings where coding time may be very limited. It was beyond the scope of this study to directly compare the identification and classification of specific vocalizations

by automated analyses versus human coding. Reduced accuracy for the automated analyses could at least partially explain the reduced effect sizes and lack of incremental validity for phonology quality.

Clinical Implications

Although automated vocal analyses are less expensive than human coding of communication samples on a per session basis, the initial cost of recording devices and access to the analysis software may be prohibitive for many clinicians. Additionally, the superior predictive validity of human-coded variables from communication samples supports the existing practice of deriving vocal variables from human coding of relatively brief communication samples. Analysis of the communicative and/or phonological quality of child vocalizations could provide clinicians with important information regarding a child's progress toward using spoken words. Although creating composites (e.g., averages) across multiple samples will improve stability of estimates and is feasible for research purposes, live coding and relatively brief communication samples would reduce the amount of coding time required and increase feasibility for assessing the quality of vocalizations. Additional investigation of such live procedures and required duration or number of sessions to produce stable estimates is warranted, as indicated above.

Conclusion

These findings support human coding of the quality of vocalizations from communication samples for young children with ASD for research purposes, despite the costs per session. These variables provided predictive value above and beyond simply measuring quantity of vocalizations. Additionally, training observers to code communication samples to derive measures of quantity and quality also provided value beyond their analogous automated variables. The LENA Research Foundation and others using their variables have not asserted that automated variables are superior to human-coded variables. However, one might argue that automated vocal variables might have lower per session cost, which can enable researchers to derive vocal variables on more participants given limited resources. Thus, our findings suggest that, if one can afford

human coding, then do so. If the investigator cannot afford human coding, then ACPU-C+V and AQV predict expressive language in low verbal children with ASD.

Acknowledgments

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9780890425596

Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., Welch, K., & Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology, 75*(4), 594–604. https://doi.org/10.1037/0022-006X.75.4.594

Book, L. A. (2009). *Early red flags for autism spectrum disorders in toddlers in the home environment* [Doctoral dissertation]. ProQuest database (UMI No. 3399180).

Camarata, S., & Yoder, P. (2002). Language transactions during development and intervention: Theoretical implications for developmental neuroscience. *International Journal of Developmental Neuroscience, 20*(3–5), 459–465. https://doi.org/10.1016/S0736-5748(02)00044-8

Cohen, J., & Cohen, P. (1984). *Applied multiple regression*. Erlbaum.

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

Enders, C. K. (2010). *Applied missing data analysis*. Guilford.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2006). *MacArthur–Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Brookes. https://doi.org/10.1037/t11538-000

Fry, D. (1966). The development of the phonological system in the normal and deaf child. In F. Smith & G. A. Miller (Eds.), *The genesis of language* (pp. 187–206). MIT Press.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences, 100*(13), 8030–8035. https://doi.org/10.1073/pnas.1332441100

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Greenwood, C. R., Carta, J. J., Walker, D., Hughes, K., & Weathers, M. (2006). Preliminary investigations of the application of the Early Communication Indicator (ECI) for infants and toddlers. *Journal of Early Intervention, 28*(3), 178–196. https://doi.org/10.1177/105381510602800306

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language, 37*(2), 229–261. https://doi.org/10.1017/S0305000909990432

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes, 16*(2–6), 287–308. https://doi.org/10.1080/01690960042000058

LENA Research Foundation. (2015). *LENA Pro*. http://www.lenafoundation.org/lena-pro

Lord, C., Risi, S., & Pickles, A. (2004). Trajectory of language development in autistic spectrum disorders. In M. L. Rice & S. F. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies*. Erlbaum.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview–Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24,* 659–685. https://doi.org/10.1007/bf02172145

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., Bishop, S., Esler, A., Hus, V., Oti, R., Richler, J., Risi, S., & Lord, C. (2009). The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39,* 1305–1320. https://doi.org/10.1007/s10803-009-0746-z

Luze, G. J., Linebarger, D. L., Greenwood, C. R., Carta, J. J., Walker, D., Leitschuh, C., & Atwater, J. B. (2001). Developing a general outcome measure of growth in the expressive communication of infants and toddlers. *School Psychology Review, 30*(3), 383–406.

McCoy, D. (2013). *Observation of social communication red flags in young children with autism spectrum disorder, developmental delay, and typical development using two observation methods* [Doctoral dissertation]. ProQuest Dissertations and Theses Global database (UMI No. 3596541 Ph.D).

McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical developmenta productivity approach. *Journal of Speech, Language, and Hearing Research, 44*(3), 670–684. https://doi.org/10.1044/1092-4388(2001/054)

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., Yoder, P., Estes, A., & Rogers, S. (2019). Validity of vocal communication and vocal complexity in young children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 50,* 224–237. https://doi.org/10.1007/s10803-019-04248-x

McLean, J., & Snyder-McLean, L. (1978). *A transactional approach to early language training*. Charles E. Merrill.

Miller, J., & Chapman, R. (2016). *Systematic Analysis of Language Transcripts* (Version 16) [Computer software]. SALT Software LLC.

Mitchell, S. K. (1979). Interobserver agreement, reliability, and generalizability of data collected in observational studies. *Psychological Bulletin, 86*(2), 376–390. https://doi.org/10.1037/0033-2909.86.2.376

Mullen, E. M. (1995). *Mullen Scales of Early Learning* (AGS ed.). Western Psychological Services.

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

Oller, D., Niyogi, P., Gray, S., Richards, J., Gilkerson, J., Xu, D., Yapane, U., & Warren, S. (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

Plumb, A. M. (2008). *Vocalizations of children with autism spectrum disorders late in the second year of life* [Doctoral dissertation]. ProQuest database (UMI No. 3348528).

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)

Rankine, J. M. (2016). *Evaluating an objective measure of language in minimally verbal autism: automated Language ENvironment Analysis (LENA) in Phelan-McDermid Syndrome* [Master's thesis]. ProQuest Dissertations and Theses Global database (UMI No. 10100456).

Rogers, S. J., Estes, A., & Yoder, P. (2013). *Intervention effects of intensity and delivery style for toddlers with ASD* (5R01MH100030) [Clinical trial]. National Institute of Mental Health.

Rvachew, S., Mattock, K., Polka, L., & Ménard, L. (2006). Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French. *The Journal of the Acoustical Society of America, 120*(4), 2250–2259. https://doi.org/10.1121/1.2266460

Sameroff, A., & Chandler, M. (1975). Reproductive risk and the continuum of caretaking casualty. In M. Horowitz, M. Hetherington, S. Scarr-Salapatek, & G. Siegel (Eds.), *Review of child development research* (pp. 187–244). University Park Press.

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

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195152968.001.0001

Sparrow, S. S., Cicchetti, D., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales–Second Edition*. AGS. https://doi.org/10.1037/t15164-000

Stoel-Gammon, C. (1991). Normal and disordered phonology in two-year-olds. *Topics in Language Disorders, 11*(4), 21–32. https://doi.org/10.1097/00011363-199111040-00005

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children*. *Journal of Child Language, 38*(1), 1–34. https://doi.org/10.1017/S0305000910000425

Swineford, L. B. (2011). *Symbol use in the home environment in toddlers suspected of having autism spectrum disorder* [Doctoral dissertation]. ProQuest Dissertations and Theses Global database (UMI No. 3502964).

Tabachnick, B., & Fidell, L. (2001). *Using multivariate statistics* (4th ed.). Allyn & Bacon.

Tager-Flusberg, H., & Joseph, R. M. (2003). Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1430), 303–314. https://doi.org/10.1098/rstb.2002.1198

Tager-Flusberg, H., & Kasari, C. (2013). 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. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (3rd ed., pp. 335–364). Wiley. https://doi.org/10.1002/9780470939345.ch12

Talbott, M. R. (2014). *Autism risk status and maternal behavior: Impacts on infant language and communication development from 6 to 36 months of age* [Doctoral dissertation]. ProQuest Dissertations and Theses Global database (UMI No. 3625962).

Tapp, J. (2003). *ProcoderDV*. Vanderbilt Kennedy Center.

Thurm, A., Lord, C., Lee, L.-C., & Newschaffer, C. (2007). Predictors of language acquisition in preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37,* 1721–1734. https://doi.org/10.1007/s10803-006-0300-1

VanDam, M., & Silbert, N. H. (2016). Fidelity of automatic speech processing for adult and child talker classifications. *PLOS ONE, 11*(8), e0160588. https://doi.org/10.1371/journal.pone.0160588

Vihman, M. (1992). Early syllables and the construction of phonology. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 393–422). York Press.

Vihman, M. (1996). *Phonological development: The origins of language in the child*. Blackwell.

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology, 108,* 1–27. https://doi.org/10.1111/bjop.12207

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language, 61*(2), 397–445. https://doi.org/10.2307/414151

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science, 25*(7), 1314–1324. https://doi.org/10.1177/0956797614531023

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*(6), 1224–1237. https://doi.org/10.1044/1092-4388(2006/088)

Wetherby, A. M., Watt, N., Morgan, L., & Shumway, S. (2007). Social communication profiles of children with autism spectrum disorders late in the second year of life. *Journal of Autism and Developmental Disorders, 37,* 960–975. https://doi.org/10.1007/s10803-006-0237-4

Woynaroski, T. (2014). *The stability and validity of automated vocal analysis in preschoolers with autism spectrum disorder in the early stages of language development* [Doctoral dissertation]. ProQuest database (UMI No. 3648771).

Woynaroski, T., Oller, D. K., Keceli-Kaysili, B., Xu, D., Richards, J. A., Gilkerson, J., Gray, S., & Yoder, P. (2017). The stability and validity of automated vocal analysis in preverbal preschoolers with autism spectrum disorder. *Autism Research, 10*(3), 508–519. https://doi.org/10.1002/aur.1667

Woynaroski, T., Yoder, P. J., Fey, M. E., & Warren, S. F. (2014). A transactional model of spoken vocabulary variation in toddlers with intellectual disabilities. *Journal of Speech, Language, and Hearing Research, 57*(5), 1754–1763. https://doi.org/10.1044/2014_JSLHR-L-13-0252

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., & Baer, C. T. (2008). *The LENA^{TM} language environment analysis system: The interpreted time segments (ITS) file*. http://www.lenafoundation.org/TechReport.aspx/ITS_File/LTR-04-2

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine, 22*(22), 3527–3541. https://doi.org/10.1002/sim.1572

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

Yoder, P., Watson, L., & Lambert, W. (2015). Value-added predictors of expressive and receptive language growth in initially nonverbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 45,* 1254–1270. https://doi.org/10.1007/s10803-014-2286-4