

Research Note

Clinician vs. Machine: Estimating Vocalizations Rates in Young Children With Developmental Disorders

Shelley L. Bredin-Oja,^a Heather Fielding,^a Kandace K. Fleming,^a and Steven F. Warren^a

Purpose: The purpose of this study was to investigate the reliability of an automated language analysis system, the Language Environment Analysis (LENA), compared with a human transcriber to determine the rate of child vocalizations during recording sessions that were significantly shorter than recommended for the automated device.

Method: Participants were 6 nonverbal male children between the ages of 28 and 46 months. Two children had autism diagnoses, 2 had Down syndrome, 1 had a chromosomal deletion, and 1 had developmental delay. Participants were recorded by the LENA digital language processor during 14 play-based interactions with a responsive adult. Rate of child vocalizations during each

of the 84 recordings was determined by both a human transcriber and the LENA software.

Results: A statistically significant difference between the 2 methods was observed for 4 of the 6 participants. Effect sizes were moderate to large. Variation in syllable structure did not explain the difference between the 2 methods. Vocalization rates from the 2 methods were highly correlated for 5 of the 6 participants.

Conclusions: Estimates of vocalization rates from nonverbal children produced by the LENA system differed from human transcription during sessions that were substantially shorter than the recommended recording length. These results confirm the recommendation of the LENA Foundation to record sessions of at least 1 hr.

Research on vocal communication has shown that young typically developing children can be reliably expected to progress from quasivowels and glottals, to sounds characterized as coos and goos, then to full vowels and marginal babbling, leading to canonical babbling (i.e., true consonants in quick succession with vowels), and finally to the emergence of first spoken words (Oller, 2000). Each stage in this predictable sequence relies at least partially on the development of the previous step. This typically occurs during the first year of life. Indeed, the ability to produce speech sounds is essential for later language success (Stoel-Gammon, 1998; Vihman, 2014). The absence of canonical babbling at 10 months old is a strong predictor of early speech delay (Oller, Eilers, Neal, & Cobo-Lewis, 1998). In addition, McCathren, Yoder, and Warren (1999) found that the amount of prelinguistic vocalizations produced by

toddlers, aged 17–34 months with developmental delays, regardless of complexity, was significantly correlated with expressive vocabulary 12 months later. McCune, Vihman, Roug-Hellichius, Delery, and Gogate (1996) found that an increase in laryngeally produced vocalizations, what the authors termed “grunts,” was followed closely by either word production or, in the case of young children with limited phonetic repertoires, use of gestures for communicative purposes. Consequently, increasing both the frequency and complexity of vocalizations is an objective of a variety of intervention approaches such as prelinguistic milieu teaching (Fey, Warren, Bredin-Oja, & Yoder, 2017) or prompts for restructuring oral muscular phonetic targets (Hayden, Eigen, Walker, & Olsen, 2010). Clinical decisions regarding the effectiveness of one intervention over another in achieving such an objective require reliable, valid measures of child vocal behavior (Law, Garrett, & Nye, 2004). However, the collection and transcription of such data are often prohibitively time consuming, thus preventing clinicians from obtaining adequate data to make informed decisions (Skahan, Watson, & Lof, 2007; Xu, Richards, & Gilkerson, 2014).

One potentially useful option for time-challenged clinicians is to employ an automated approach to data collection and analysis such as the Language Environment

^aThe Schiefelbusch Institute of Life Span Studies, University of Kansas, Lawrence

Correspondence to Shelley L. Bredin-Oja: sbredin-oja@ku.edu

Editor-in-Chief: Julie Barkmeier-Kraemer

Editor: Cynthia Cress

Received February 2, 2017

Revision received July 7, 2017

Accepted March 19, 2018

https://doi.org/10.1044/2018_AJSLP-17-0016

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Analysis (LENA) system (Gilkerson & Richards, 2009). The LENA system uses a small, lightweight digital language processor (DLP) placed in the front pocket of specially designed clothing that is worn by a young child. This allows the unobtrusive audio recording of up to 12 hr of verbal interaction in a single day. After making a recording, the DLP is connected to a computer, and specialized software processes it using an algorithm that extracts features from the audio file, segments the sounds, and then classifies the segments into eight major categories comprising male and female adult speakers, key child, other child, overlap, noise, TV/electronic sounds, and silence (Ford, Baer, Xu, Yapanel, & Gray, 2009). Overlap includes sounds from more than one source, for example, two people talking simultaneously or a child vocalizing during other noise such as toys banging loudly together. If the algorithm cannot reliably classify the audio signal into a single category, that segment of the file is classified as overlap (Ford et al., 2009). The algorithm provides an estimate of the number of adult words directed to or overheard by the child, the amount of child vocalizations, and the number of conversational turns that occur in the child's environment over the length of the recording. Importantly, segments classified as overlap do not contribute to these estimates. Child vocalizations are defined as speech-related sounds including coos, raspberries, babbling, and words. The LENA algorithm filters out vegetative sounds and fixed signals such as sneezing, coughing, laughing, and crying. Each child vocalization that is separated by at least a 300-ms pause is tallied, and the frequency is provided in reports based on user preference (e.g., daily, hourly, 5-min blocks). In addition, the user may export audio files to a .wav format for transcription.

Differences between the two methods consist of more than only an investment of time; transcription by hand is a much lengthier process. A human transcriber can reliably distinguish speech-related sounds from vegetative sounds and fixed signals just as the LENA system can. However, human transcribers may be more adept at recognizing a child vocalization in the presence of overlapping sounds such as an adult talking or a loud noise from a toy, sounds that the LENA would typically "throw out" as overlap. On the other hand, discerning a break in phonation that is only 300 ms may be more of a challenge for the human ear.

Although there are distinct reasons the two methods may provide different estimates, it is important to note that the estimates generated by the LENA algorithm become more accurate the longer the recording period because individual mistakes in classifications tend to cancel each other out over time. The system was designed and normed for use in a 12-hr-long spontaneous adult and child speech environment (Xu, Yapanel, & Gray, 2009). A technical report released by the LENA Foundation showed high agreement (i.e., 82% for adult segments and 76% for child segments) between human-transcribed and LENA-based segmentation of 70 hr of transcription (Xu et al., 2009). In addition, the reliability of the LENA system over daylong

recordings has been empirically established in a number of studies (Oller et al., 2010; Warren et al., 2010; Xu et al., 2014). However, it has also been reported by the developers that accurate information may be generated if the device is worn by a child for a minimum of 1 hr (Xu et al., 2009).

The LENA system has been used to study a variety of populations, including typically developing children and children with a variety of developmental challenges including Down syndrome, hearing loss, autism, environmental risk, and severe prematurity (e.g., Caskey, Stephens, Tucker, & Vohr, 2011; Suskind et al., 2016; Thiemann-Bourque, Warren, Brady, Gilkerson, & Richards, 2014; VanDam et al., 2015; Warren et al., 2010; Zimmerman et al., 2009). In addition, the LENA system has been deployed in a range of environments including children's homes, classrooms, and even neonatal care units (e.g., Caskey, Stephens, Tucker, & Vohr, 2014; Dykstra et al., 2013; Gilkerson & Richards, 2008). The automated nature of the LENA system for measuring child vocal behavior offers a potentially valuable tool to clinicians who often do not have sufficient time for the labor-intensive collection and analysis of child vocalizations (Skahan et al., 2007; Xu et al., 2014). For example, obvious clinical uses of the LENA system may be to automatically measure pretreatment child vocalizations and posttreatment effects of therapy.

The purpose of this study was to examine the reliability of a much shorter data collection period than the LENA developers suggest. Specifically, we examined the extent to which the LENA system may provide similar estimates for the number of child vocalizations compared with a human transcriber during recording sessions averaging only 23 min. If shorter samples are sufficiently reliable, this could enhance the utility of the LENA system to gather and analyze child speech data in more time-limited contexts commonly encountered in clinical practice.

Method

Participants

Six boys, aged 31–46 months ($M = 36$ months), with delayed communication development participated in 14 individual sessions each over the course of 2 months as part of another study (Bredin-Oja, Fielding, & Warren, 2016). Briefly, that study investigated the effect of a speech-generating device on the rate and complexity of child vocalizations (Bredin-Oja, Fielding, & Warren, 2018). Each child had a diagnosis known to negatively impact speech and language development. Children were recruited from local Part C service providers (see Table 1 for participant characteristics).

Procedure

During each session, the child wore the DLP in the front chest pocket of a specially designed vest while he interacted with a highly responsive adult who followed the

Table 1. Participant characteristics.

Participant ID	Sex	Age (months)	Diagnosis	Number of words
1	M	31	Autism spectrum disorder	0
2	M	44	16p11.2 Deletion	0
3	M	46	Autism spectrum disorder	0
4	M	28	Developmental delay	1 (bye)
5	M	28	Down syndrome	0
6	M	39	Down syndrome	0

Note. M = male.

principles and procedures of prelinguistic milieu teaching (Fey et al., 2017). That is, the adult arranged the environment to create opportunities for the child to communicate, such as placing toys in view but out of reach; followed the child's attentional lead; prompted the child to initiate a request by waiting expectantly or asking open-ended questions such as "What do you want?"; and then contingently responded by providing the requested object or action. In seven of the sessions, the child had access to and was required to use a speech-generating device to communicate. In the remaining seven sessions, the child did not have access to the speech-generating device. Rather, the child was prompted to intentionally communicate through prelinguistic means of coordinated gaze and gestures. Vocalizations were not prompted, as this was the outcome variable of interest in the other study. Specifically, we asked what impact a speech-generating device has on the rate and complexity of child vocalizations (Bredin-Oja et al., 2018). During each session, the child was prompted 20 times to request a toy or an action at a rate of approximately one prompt per minute—the rate recommended by Fey and colleagues (2017). Sessions lasted between 17 and 31 min, with a mean length of 23 min.

After each session, the LENA Pro software (Oller et al., 2010; Xu et al., 2009; Xu, Yapanel, Gray, & Baer, 2008) was used to process and analyze the audio recording. In addition, Child Language Analysis software (MacWhinney, 2000) was used to export the audio recording and generate a .wav file for transcription. The second author, a doctoral student in child language with previous training and experience in phonetic transcription, transcribed all child vocalizations that had clear vocal fold vibration and were not vegetative noises such as burps, hiccups, and sounds of exertion, or fixed signals such as sighs, ingressive vocalizations, laughter, whines, or cries. The occurrence of each vocalization was marked in minutes and seconds from the beginning of the vocalization. Vocalizations were segmented by a breath, a 0.5-s pause, or a descending intonation pattern followed by an obvious break in phonation. The pause length differed from the LENA software because 300 ms proved to be too difficult for the human transcribers to discern reliably. Complexity of child vocalizations was determined according to the syllable structure level (Paul & Jennings, 1992). Briefly, the syllable structure level comprises three levels: Level 1 utterances are composed of a voiced vowel, a voiced syllabic consonant, or a consonant–vowel containing a

glottal stop or glide; Level 2 utterances are composed of a vowel–consonant or a consonant–vowel–consonant with a single consonant type other than a glottal stop or glide, disregarding voicing differences; and Level 3 utterances are composed of syllables with two or more different consonant types, disregarding voicing differences. Sessions varied in length; therefore, the rate of vocalizations per minute was used as the comparison between the LENA software and the human transcriber.

Reliability

The first author transcribed 29% (4/14) of each child's files to calculate the reliability of the primary transcriber's coding. Reliability was calculated using line-by-line exact agreement and a confusion matrix, a cross-tabulation of ordered scores assigned by each rater with agreements along the diagonal. Scores farther from the diagonal indicate more extreme disagreement, whereas scores closer to the diagonal are more similar (Kohavi & Provost, 1998). Vocalizations were considered to be the same event if they occurred within 1 s of each other; vocalizations that differed by greater than 1 s were considered different events and therefore were counted as a disagreement. The number of agreements divided by the number of agreements plus disagreements was calculated as the proportion of agreement. Reliability (proportion of agreement) for all six children ranged from .76 to .90 ($M = .83$). Reliability for assigning the syllable structure level ranged from .78 to 1.0 ($M = .90$). Both transcribers were blind to the LENA counts for a single child until all 14 sessions had been transcribed.

Results

Individual paired-samples two-tailed *t* tests revealed a statistically significant difference in terms of vocalization counts between the LENA system and the human transcriber for four of the six children. Effect sizes were moderate to large for all children (see Table 2 for individual results). The actual differences between the human transcriber and the LENA system were relatively small, ranging from 0.29 to 1.35 vocalizations per minute. This equates to a proportional difference ranging from 0.13 to 0.30 vocalizations per minute, with an average of 0.21 or a

Table 2. Individual paired-samples *t* tests.

ID	LENA rate	Transcriber rate	Mean difference	SD of difference	SEM	95% CI		<i>t</i>	df	Sig. (two-tailed)	Effect size (<i>d</i>)
						Lower	Upper				
1	3.52	4.06	-0.54	0.806	0.215	-1.002	-0.713	-2.492	13	.027	0.668
2	4.49	5.77	-1.28	1.808	0.483	-2.339	-0.241	-2.660	13	.020	0.726
3	1.27	0.98	0.29	0.631	0.169	-0.073	0.655	1.727	13	.108	0.517
4	4.25	3.70	0.55	0.954	0.255	-0.002	1.099	2.152	13	.051	0.605
5	4.11	4.89	-0.78	1.307	0.349	-1.529	0.020	-2.218	13	.045	0.613
6	5.78	4.43	1.35	0.661	0.177	0.967	1.730	7.637	13	.000	2.138

Note. SEM = standard error of measurement; CI = confidence interval; LENA = Language Environment Analysis.

21% difference in rate between counts produced by each method.

It may be that either the LENA system or the human transcriber more accurately counts vocalizations that are canonical than vocalizations comprising only vowels. To determine whether the complexity of vocalizations accounted for differences between the two methods, we sought to compare vocal complexity in sessions showing high agreement with sessions showing poor agreement. The LENA does not differentiate between noncanonical vocalizations, canonical vocalizations, and even true words in its child vocalization count. However, as a part of the human transcription procedure, each vocalization was assigned a syllable structure level (Paul & Jennings, 1992). Unfortunately, there was so little variation across all sessions within each child that this analysis did not yield any useful information.

To further examine the similarity in vocalization rates for each child, scores obtained across the 14 sessions using each scoring system were correlated within each child (see Table 3 for individual results). For five of the six children, the correlations between the LENA system and the human transcriber were moderate to high, which indicates that, although the two systems do not determine the same rate, they are ordering the amount of communication similarly across observations. Furthermore, this association between the two types of scores holds between children, in that the children with the highest rates of vocalizations as counted by the LENA also had the highest rates according to the human transcriber.

Figure 1 illustrates how different the human transcriber's counts are from the LENA system for one child (i.e., Child 2)

Table 3. Paired-samples correlations between LENA count and human transcriber count.

Participant ID	<i>N</i>	Correlation	Significance
1	14	.879	.000
2	14	-.019	.949
3	14	.628	.016
4	14	.765	.001
5	14	.572	.033
6	14	.959	.000

Note. LENA = Language Environment Analysis.

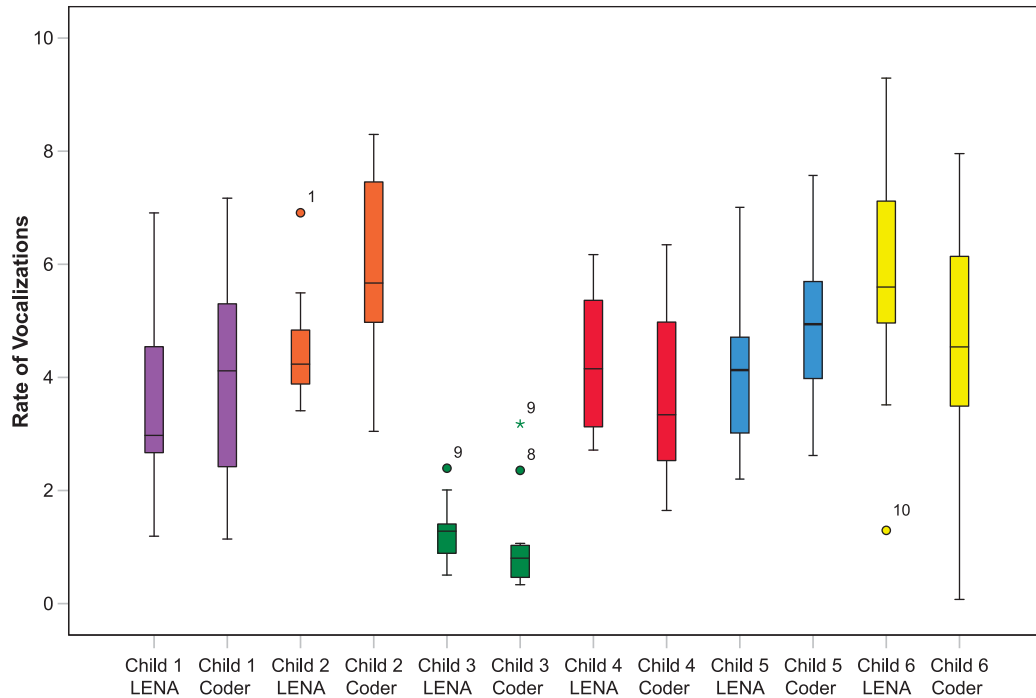
whose correlation between scores was near zero. Environmental factors may offer an explanation for the differences. During each of this particular child's 14 sessions, two other siblings, one close in age and one infant, were in close proximity to the key child. It is probable that some of the key child's vocalizations occurred while another child was vocalizing and were thus categorized as overlap by the LENA system and therefore not included in the tally of child vocalizations. In contrast, the human transcriber could more easily discriminate between the key child's vocalizations and other overlapping sounds or vocalizations, resulting in differences in the number of vocalizations counted. Sessions for the other five children did not have a similar environment; there were either no other children in the home, other children remained in a different part of the home, or sessions were conducted in a university laboratory setting.

Discussion

Frequent and complex vocalizations are fundamental for the acquisition of spoken language (McCathren et al., 1999; McCune et al., 1996; Oller et al., 1998; Stoel-Gammon, 1998; Vihman, 2014) and thus are often targeted by clinicians who work with children with communication delays. Yet, insufficient data collection and analysis may be a barrier to informed clinical decisions regarding the effectiveness of intervention (Skahan et al., 2007). The automation of the LENA system is an appealing solution to this problem. This study examined the reliability of the LENA system to estimate a child's vocalizations during play-based sessions that were much shorter than the recording length recommended by the LENA Foundation (Xu et al., 2009).

Individual *t* tests revealed a statistically significant difference between rates of vocalizations counted by the LENA system and those of a human transcriber for four of the six children. Further analysis regarding the complexity of vocalizations did not explain differences between the two methods because children rarely varied in the complexity of their vocalizations across sessions. Correlations between rates obtained by the two methods were moderate to high for five of the six children, indicating similar ordering of rates despite differences in actual rates. Environmental factors likely resulted in a nonsignificant correlation for

Figure 1. Boxplot distributions of LENA count and human transcriber count across 14 sessions. The dots indicate sessions with outlying rates of vocalizations, whereas the asterisks are extreme outliers that are three times the interquartile range from the median. For example, Session 9 for Child 3 had an outlying rate of vocalizations for both the LENA system and the human transcriber. LENA = Language Environment Analysis.



one child. Thus, the two methods captured similar patterns in the rate of vocalizations. Nevertheless, the two methods clearly gave different counts for four of the six children. These differences are magnified by moderate to large effect sizes. The LENA system is a useful tool for quantifying the language environment of children over all day recordings (Oller et al., 2010); however, results from this small, single-case–design study do not indicate that results from the LENA system can be relied upon to give accurate vocalization rates of nonverbal children for sessions as short as 20–25 min.

Study Limitations

This study investigated the reliability of the LENA system compared with human transcription for sessions that were, on average, 23 min in length. We did not compare frequency counts between the two methods for sessions that are longer than this but still shorter than the daylong recordings that have been validated in the literature. It must also be noted that, despite the recommendation of the LENA Foundation that recording sessions be at least 1 hr in length, the accuracy of this length of time has not been validated for children with developmental disabilities who are at the prelinguistic stage of development. Further research is needed to validate LENA data from sessions that are less than daylong recordings.

Clinical Implications

The frequency of vocalizations, regardless of complexity, has been found to predict later communication ability (McCathren et al., 1999; McCune et al., 1996; Vihman, 2014). Therefore, intervention should target increasing the rate of all vocalizations for prelinguistic children showing evidence of a developmental language delay. However, adequately documenting the impact of targeting this behavior requires a labor-intensive, time-consuming effort. Five participants had highly correlated rates of vocalization between the LENA and the human transcriber. For these participants, a similar rate of vocalizations was captured by both the LENA and the human transcriber during the same session (e.g., for Participant 6 during Session 2, the LENA reported 9.29 vocalizations per minute and the human transcriber reported 7.96; during Session 14, the LENA reported 5.42 and the human transcriber reported 4.57). Therefore, the LENA captured similar patterns to the human transcriber within each session. Although the LENA may have difficulty providing an exact number of vocalizations, it could be useful to capture general increases or decreases in frequency of vocalizations during shorter recording sessions. This could be useful for time-pressed clinicians who wish to determine if intervention affects rates of vocalizations rather than precise counts of vocalizations. Although our findings do not validate relying on the automated LENA system for precise vocalization

counts during sessions as short as 20–25 min, the LENA system may still be a valuable source to determine general trends in rate or frequency of vocalizations.

In addition, the LENA system records the child's vocalizations in a manner that is unobtrusive and highly portable; the DLP fits easily into a pocket of a shirt or vest worn by the child, which eliminates the need to ensure that the child is always near a microphone. The LENA software also provides estimates of child vocalizations in 5-min increments. This enables a clinician to easily record an intervention session and, after viewing the LENA report, identify which segments of the recording have frequent vocalizations as well as segments that have few or none. The clinician could then choose which segments to listen to and code these by hand without having to listen to the entire recording if exact counts are required. In short, the LENA system is a clinically useful tool that can be used in a variety of contexts to help clinicians track a child's vocalizations.

Acknowledgments

This work was supported by a National Institute of Child Health and Development grant (P30 HD002528, awarded to John Colombo, PI) and by a National Institute of Deafness and Other Communication Disorders training grant (T32 DC000052, awarded to Mabel Rice, PI).

References

- Bredin-Oja, S. L., Fielding, H., & Warren, S. F. (2016, November 18). *Using LENA to measure effects of a speech generating device on child vocalizations*. Poster presented at the annual convention of the American Speech-Language-Hearing Association, Philadelphia, PA.
- Bredin-Oja, S. L., Fielding, H., & Warren, S. F. (2018). *Effect of a speech generating device on child vocalizations*. Manuscript in preparation.
- Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2011). Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics*, *128*, 910–916. <https://doi.org/10.1542/peds.2011-0609>
- Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2014). Adult talk in the NICU with preterm infants and developmental outcomes. *Pediatrics*, *133*, e578–e584. <https://doi.org/10.1542/peds.2013-0104>
- 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*, 582–594. <https://doi.org/10.1177/1362361312446206>
- Fey, M. E., Warren, S. F., Bredin-Oja, S. L., & Yoder, P. J. (2017). Responsivity education/prelinguistic milieu teaching. In R. McCauley, M. E. Fey, & R. B. Gillam (Eds.), *Treatment of language disorders in children* (2nd ed., pp. 57–86). Baltimore, MD: Brookes.
- Ford, M., Baer, C. T., Xu, D., Yapanel, U., & Gray, S. (2009). *The LENA language environment analysis system: Audio specifications of the DLP-0121* (LENA Foundation Technical Report LTR-03-2). Retrieved from http://lena.org/wp-content/uploads/2016/07/LTR-03-2_Audio_Specifications.pdf
- Gilkerson, J., & Richards, J. A. (2008). *The LENA natural language study* (LENA Foundation Technical Report LTR-02-2). Retrieved from http://lena.org/wp-content/uploads/2016/07/LTR-02-2_Natural_Language_Study.pdf
- Gilkerson, J., & Richards, J. A. (2009). *The power of talk: Impact of adult talk, conversational turns, and TV during the critical 0–4 years of child development* (LENA Foundation Technical Report LTR-01-2). Retrieved from http://lena.org/wp-content/uploads/2016/07/LTR-01-2_PowerOfTalk.pdf
- Hayden, D., Eigen, J., Walker, A., & Olsen, L. (2010). PROMPT: A tactually grounded model. In L. Williams, S. McLeod, & R. McCauley (Eds.), *Interventions for speech sound disorders in children* (pp. 453–474). Baltimore, MD: Brookes.
- Kohavi, R., & Provost, F. (1998). Confusion matrix. *Machine Learning*, *30*, 271–274. <https://doi.org/10.1023/A:1017181826899>
- Law, J., Garrett, Z., & Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *47*, 924–943. [https://doi.org/10.1044/1092-4388\(2004\)069](https://doi.org/10.1044/1092-4388(2004)069)
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Erlbaum.
- McCathren, R. B., Yoder, P. J., & Warren, S. F. (1999). The relationship between prelinguistic vocalizations and later expressive vocabulary in young children with developmental delay. *Journal of Speech, Language, and Hearing Research*, *42*, 915–924. <https://doi.org/10.1044/jslhr.4204.915>
- McCune, L., Vihman, M. M., Roug-Hellichius, L., Delery, D. B., & Gogate, L. (1996). Grunt communication in human infants (homo sapiens). *Journal of Comparative Psychology*, *110*, 27–37. <https://doi.org/10.1037/0735-7036.110.1.27>
- Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Erlbaum.
- Oller, D. K., Eilers, R. E., Neal, A. R., & Cobo-Lewis, A. B. (1998). Late onset canonical babbling: A possible early marker of abnormal development. *American Journal on Mental Retardation*, *103*, 249–263. [https://doi.org/10.1352/0895-8017\(1998\)103<0249:LOCBAP>2.0.CO;2](https://doi.org/10.1352/0895-8017(1998)103<0249:LOCBAP>2.0.CO;2)
- Oller, D. K., Niyogi, P., Gray, S., Richards, J., Gilkerson, J., Xu, D., . . . Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences, U.S.A.*, *107*, 13354–13359. <https://doi.org/10.1073/pnas.1003882107>
- Paul, R., & Jennings, P. (1992). Phonological behavior in toddlers with slow expressive language development. *Journal of Speech and Hearing Research*, *35*, 99–107. <https://doi.org/10.1044/jshr.3501.99>
- Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology*, *16*, 246–259. [https://doi.org/10.1044/1058-0360\(2007\)029](https://doi.org/10.1044/1058-0360(2007)029)
- Stoel-Gammon, C. (1998). The role of babbling and phonology in early linguistic development. In A. M. Wetherby, S. F. Warren, & J. Reichle (Eds.), *Transitions in prelinguistic communication* (Vol. 7, pp. 87–110). Baltimore, MD: Brookes.
- Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., . . . Levine, S. C. (2016). A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study. *Journal of Child Language*, *43*, 366–406. <https://doi.org/10.1017/S0305000915000033>

-
- Thiemann-Bourque, K. S., Warren, S. F., Brady, N., Gilkerson, J., & Richards, J. A.** (2014). Vocal interaction between children with down syndrome and their parents. *American Journal of Speech-Language Pathology, 23*, 474–485. https://doi.org/10.1044/2014_AJSLP-12-0010
- VanDam, M., Oller, D. K., Ambrose, S. E., Gray, S., Richards, J. A., Xu, D., . . . Moeller, M. P.** (2015). Automated vocal analysis of children with hearing loss and their typical and atypical peers. *Ear and Hearing, 36*, e146–e152. <https://doi.org/10.1097/AUD.0000000000000138>
- Vihman, M. M.** (2014). *Phonological development: The first two years* (2nd ed.). Boston, MA: Blackwell.
- Warren, S. F., Gilkerson, J., Richards, J. A., Oller, D. K., Xu, D., Yapanel, U., & Gray, S.** (2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of Autism and Developmental Disorders, 40*, 555–569. <https://doi.org/10.1007/s10803-009-0902-5>
- 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*, 1638–1650. https://doi.org/10.1044/2014_JSLHR-S-13-0037
- Xu, D., Yapanel, U., & Gray, S.** (2009). *Reliability of the LENA language environment analysis system in young children's natural home environment* (LENA Research Foundation Technical Report LTR-05-2). Retrieved from http://lena.org/wp-content/uploads/2016/07/LTR-05-2_Reliability.pdf
- Xu, D., Yapanel, U., Gray, S., & Baer, C. T.** (2008). *The LENA language environment analysis system: The interpreted time segments (ITS) file* (LENA Foundation Technical Report LTR-04-2). Retrieved from http://lena.org/wp-content/uploads/2016/07/LTR-04-2_ITS_File.pdf
- Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U.** (2009). Teaching by listening: The importance of adult–child conversations to language development. *Pediatrics, 124*, 342–349. <https://doi.org/10.1542/peds.2008-2267>