HHS Public Access
Author manuscript
Dev Rev. Author manuscript; available in PMC 2020 September 01.

Published in final edited form as:
Dev Rev. 2020 September ; 57: . doi:10.1016/j.dr.2020.100921.

A meta-analysis of the predictability of LENA™ automated measures for child language development

Yuanyuan Wang[1,*], **Rondeline Williams**[1], **Laura Dilley**[2], **Derek M. Houston**[1,3]

[1]Department of Otolaryngology–Head & Neck Surgery, The Ohio State University, 915 Olentangy River Road # 4000, Columbus, OH

[2]Department of Communicative Sciences & Disorders, Michigan State University, East Lansing, MI 48824

[3]Nationwide Children's Hospital, Columbus, OH, 700 Children's Drive, Columbus, OH 43205

Abstract

Early language environment plays a critical role in child language development. The Language ENvironment Analysis (LENA™) system allows researchers and clinicians to collect daylong recordings and obtain automated measures to characterize a child's language environment. This meta-analysis evaluates the predictability of LENA's automated measures for language skills in young children. We systematically searched reports for associations between LENA's automated measures, specifically, adult word count (AWC), conversational turn count (CTC), and child vocalization count (CVC), and language skills in children younger than 48 months. Using robust variance estimation, we calculated weighted mean effect sizes and conducted moderator analyses exploring the factors that might affect this relationship. The results revealed an overall medium effect size for the correlation between LENA's automated measures and language skills. This relationship was largely consistent regardless of child developmental status, publication status, language assessment modality and method, or the age at which the LENA recording was taken; however, the effect was weakly moderated by the gap between LENA recordings and language measures taken. Among the three measures, there were medium associations between CTC and CVC and language, whereas there was a small-to-medium association between AWC and language. These findings extend beyond validation work conducted by the LENA Research Foundation and suggest certain predictive strength of LENA's automated measures for child language. We discussed possible mechanisms underlying the observed associations, as well as the theoretical, methodological, and clinical implications of these findings.

Keywords

LENA; AWC; CTC; CVC; predictability; child language

Early language environment plays a critical role in child language and cognitive development that is related to later personal, academic, and social achievements (e.g.,

*Address for correspondence: Yuanyuan Wang, Department of Otolaryngology-Head and Neck Surgery, The Ohio State University, 915 Olentangy River Rd., Columbus, OH 43212; Yuanyuan.Wang@osumc.edu.

Gilkerson et al., 2018; Hart & Risley, 1995; Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002; Topping, Dekhinet, & Zeedyk, 2011). In the landmark longitudinal study, Hart and Risley (1995) examined the amount of language that children heard from 42 Kansas families of various socioeconomic status. They recorded monthly hour-long interactions between caregivers and their children from age 7 months to 3 years, and calculated the number of words caregivers produced. Based on these calculation, it was estimated that by 4 years of age, children from low-income families were exposed to 30 million fewer words (known as the 30-million Word Gap) than children from professional families. More importantly, the amount of caregiver speech to children also significantly predicted child language and cognitive development later in life. Much subsequent research has extended this work and examined the properties of early language environment and their relationships developmental outcomes (Greenwood, Thiemann-Bourque, Walker, Buzhardt, & Gilkerson, 2011; Hoff & Naigles, 2002; Romeo et al., 2018; Rowe, 2012; Weisleder & Fernald, 2013; Weizman & Snow, 2001). Collectively, this body of work provides strong evidence that early language experience is related to various aspects of development.

While naturalistic language input recorded from the home environment provides spontaneous speech material, the form of data collection and analysis used in Hart and Risley (1995)'s and subsequent work required extensive time. For example, In Hart and Risley (1995)'s work, 1-hour recording samples required 8 to 10 hours' of human transcription, which was very labor-intensive and time-consuming. The costs and challenges associated with methodologies for analyzing a massive amount of language input have apparently been proven a major barrier for conducting large-scale studies examining the naturalistic language environment. Therefore, most studies examining child language experience thus far were conducted in laboratory settings, with short audiovisual recordings and often with experimenters present moderating the recording sessions. This form of data collection, while practical and valuable, may not be representative of children's authentic language environment and thus may not be generalizable. Indeed, recently, Sperry, Sperry, and Miller (2019) criticized that Hart and Risley (1995)'s study only included a subset of the total number of words spoken to the children, which did not reflect children's full language environment. Therefore, gathering and analyzing naturalistic speech samples from the home environment for a longer period than was previously practical would be an important next step for the field.

With the development of automatic speech processing (ASP) technologies, new tools were designed and invented that allowed for automated analysis of speech data. The Language Environment Analysis (LENA™, the LENA Research Foundation, Boulder, CO) system is the first ASP device that was designed to automatically analyze speech occurring in the naturalistic home environment (Christakis et al., 2009; Xu, Yapanel, & Gray, 2009; Zimmerman et al., 2009). The initial goal of the LENA Research Foundation into developing the LENA system was to provide an automatic estimate of Hart and Risley's indices of the amount of language occurring in the home environment, and thus informing and guiding parents and clinicians seeking solutions to close the Word Gap. Since the advent of the LENA system, researchers and clinicians have been using it to measure various aspects of the child language environment, including the quantity of speech and interactions with caregivers in the home. Moreover, LENA has also been used to examine whether these measures contribute to explaining variabilities in children's linguistic and cognitive outcomes (Adams et al., 2018; Ambrose, VanDam, & Moeller, 2014; Gilkerson et al., 2018; VanDam, Ambrose, & Moeller, 2012; Weisleder & Fernald, 2013). The primary purpose of the present study was to explore this latter line of work and examine the predictive validity of LENA's automatically generated metrics for child language skills using systematic review and meta-analysis methodology.

The LENA system

The LENA system consists of a digital recorder and software that automatically processes the audio. The LENA recorder is a compact wearable digital language processor (DLP) which collects daylong recordings of a child's early language environment. It measures approximately 3–3/8" x 2–3/16" x 1/2", and weighs about three ounces. The LENA DLP is secured in a pocket on a specially designed vest or clothes worn by the target child; it allows for up to 16 hours of continuous recording of speech data collected in the vicinity of the target child. Once the data collection is completed, the LENA DLP is connected to a computer; the LENA software automatically uploads and analyzes the auditory data using a series of iterative modeling algorithms developed by the LENA Research Foundation. It then segments the daylong recordings based on acoustic energy and generates three measures that are related to Hart and Risley's indices. These measures include adult word count (AWC; the total number of adult words spoken near the target child who wears the device), conversational turn count (CTC; the total number of conversational interactions the child engages in with an adult in which one speaker initiates and the other responds within five seconds), and child vocalization count (CVC; the total number of speech-like utterances produced by child). The LENA software also generates other classifications, including overlapping speech, TV and media, background noise, and silence to characterize different aspects of the child auditory environment.

In this meta-analysis, we focused on measures of AWC, CTC, and CVC out of both theoretical and methodological considerations. From a theoretical perspective, these measures are the most relevant to Hart and Risley's indices to characterize the child language environment. One may argue that CVC is a measure of child's own linguistic productions, but not necessarily a measure of child language environment. We included CVC out of two considerations. First, according to auditory feedback models, hearing one's own voice is required for vocal learning as it allows for an evaluation of the auditory feedback relative to the adult template (Brainard & Doupe, 2000). Second, research has shown that child vocal development reflects the adult language model and is driven by interaction with other social entities (Moeller et al., 2007). Measures of AWC, CTC, and CVC may not be independent of each other, as both theoretical and empirical evidence indicates a dynamic, reciprocal, and contingent vocal interactions between caregivers and their infants (Goldstein, King, & West, 2003; Pretzer, Lopez, Walle, & Warlaumont, 2019; Sameroff, 1975). That is, a larger number of AWC may accompany larger numbers of CVC and CTC. From a methodological perspective, there are not enough studies that have reported data on other measures to be included in the meta-analysis. According to recent reviews on the research using the LENA system, research thus far has focused primarily on AWC along with CTC and CVC, whereas research that has included other automated measures is very limited (Ganek & Eriks-Brophy, 2018; Greenwood, Schnitz, Irvin, Tsai, & Carta, 2018). For example, in a most recent review article on the use of the LENA system, Greenwood et al. (2018) showed that whereas 39 reports included AWC, 32 included CTC, and 31 included CVC, only 4 reports examined TV and electronic sounds, and 2 included analyses of background noise.

LENA accuracy and reliability

Although LENA removes the most central barrier in analyzing massive-scale naturalistic speech and has been adopted in a variety of settings by researchers, clinicians, educators, and parents, questions remain about the validity of the LENA system in quantifying the language environment. Before we discuss LENA's predictability strength, we first overview the research that has assessed LENA's validity (Gilkerson, Coulter, & Richards, 2008; VanDam & Silbert, 2016; Xu et al., 2009).

The first accuracy and reliability of LENA's automated analyses were assessed by the LENA Research Foundation through comparisons between LENA's automated measures to human transcription. They collected 70 12-hour long audio files from children 2 to 36 months and transcribed one hour of each file. These data were also included as part of the Natural Language Study (NLS), the LENA Research Foundation's normative study (Gilkerson et al., 2008; Gilkerson et al., 2018). The results showed that the segmentation agreement percentages between LENA software and human transcribers were 82%, 76%, 71%, and 76% for adult, child, TV, and other speech, respectively (Xu et al., 2009). Moreover, the Pearson correlation between LENA and human-based AWC estimates was high, $r = .92$; however, correlations between LENA output and human-based transcription for the measures of CTC and CVC were not reported.

Subsequent research has also compared LENA's automated measures with human transcriptions (Busch, Sangen, Vanpoucke, & van Wieringen, 2018; Gilkerson et al., 2015; Soderstrom & Wittebolle, 2013; VanDam & Silbert, 2016). For example, Soderstrom and Wittebolle (2013) examined AWC and CVC from 183 5-minute speech samples from 11–20-month-old English-speaking children recorded at home and daycare settings. The Pearson correlation coefficients between the LENA and human-coded data were $r = .76$ and $r = .69$ for AWC and CVC, respectively. Gilkerson et al. (2015) assessed LENA AWC and CTC estimates for the Chinese Shanghai dialect and Mandarin languages. They selected 5-minute speech samples from 22 children between the ages of 3 and 23 months. The results showed a correlation coefficient of $r = .72$ between LENA generated and human transcribed measures for AWC. However, the Pearson correlation coefficient for CTC was very low, $r = .22$. They explained that the low correlation coefficient for CTC was mainly due to the three outliers that they identified; after excluding the three outliers, the results yielded a correlation of $r = .72$ for CTC. Moreover, Busch et al. (2018) selected samples from 8 daylong recordings of 6 Dutch-speaking children between 2 and 5 years of age. The correlation coefficients were $r = .87$ for AWC, $r = .52$ for CTC, and $r = .77$ for CTC. Recently, Cristia, Bulgarelli, and Bergelson (2020) conducted a meta-analysis comparing LENA's automated measures to human transcription. The findings showed a mean correlation coefficient of $r = .79$ for AWC (based on 13 reports), $r = .36$ for CTC (based on 6 reports; note that the low correlation for CVC was largely pulled by the three outliers in Gilkerson et al. (2015)'s study), and $r = .77$ for CVC (based on 5 reports). In sum, the above-mentioned small number of studies that have compared LENA's automated measures with human transcriptions mainly focused on AWC, with less emphasis on CTC, and CVC. Findings from this research suggested a relatively reliable correlation between LENA's automated measures and human transcription for AWC and CVC, but less reliable correlation for CTC.

In spite of these findings, it should be noted that most validation work tended to assume that human coders were often accurate, and any inconsistency between LENA automated output and human transcriptions was due to LENA's error. However, we should acknowledge that humans also make errors in transcription (Stolcke & Droppo, 2017; Xiong et al., 2017). For example, Xiong et al. (2017) measured human error rate in a speech recognition task. They found that the professional transcribers had error rates of 5.9% and 11.3%, which was slightly higher than the automated speech transcription system that had the error rates of 5.8% and 11.0%. Among the studies which compared LENA automated and human coded measures, very few have reported inter-coder reliability for human coders. For example, Soderstrom and Wittebolle (2013) compared two human coders' transcriptions and reported an inter-coder agreement of 76% based on the results from 216 5-minute segments. Furthermore, validation studies have been limited to the examination of a small sample of speech out of practical considerations; therefore, it is still unclear whether LENA's reliability would change as a function of sampling size/duration, that is, it is possible that the correlations between LENA's automated measures and human transcriptions will be higher when longer samples of LENA data are transcribed and compared. Therefore, the majority of validation studies, which transcribed 5- or 10-minutes randomly sampled segments, may not have best reflected LENA's accuracy. Regardless, it is important to continue evaluating LENA's accuracy and reliability using a variety of methodologies, sampling methods, and analyses. Results from this line of work are likely to improve the LENA algorithms and procedures to generate more accurate outputs in the future.

Using LENA's automated measures for predicting language outcomes

Since the introduction of the LENA system, a growing body of work has examined whether LENA's automatically generated metrics, (e.g., AWC, CTC, and CVC), contributed to explaining variability in child language and cognitive development. These studies involved infants and children with diverse developmental status, including typically-developing (TD) infants (Gilkerson et al., 2018; Greenwood et al., 2011), preterm infants (Caskey, Stephens, Tucker, & Vohr, 2014), infants with autism spectrum disorder (ASD) (Patterson, 2010), and infants with hearing loss (HL) (Ambrose et al., 2014; VanDam et al., 2012), and from different language backgrounds, including infants learning English (Gilkerson et al., 2018; Greenwood et al., 2011), Finnish (Elo, 2016), and Mandarin (Xu, Zhang, Mao, Xin, & Xiao, 2012).

For example, Ambrose et al. (2014) collected daylong recordings of the auditory environment from 28 children with mild-to-severe hearing loss within 6 months of their second birthday. On average, the children received approximately 1400 AWC per hour and participated in 60 conversational blocks per hour, although there was a considerable within-group variability of the AWC and CTC. Furthermore, they showed that CTC, but not AWC, significantly predicted children's communication outcomes measured at 2 and 3 years of age. Similarly, Caskey et al. (2014) collected recordings from 36 preterm infants in the NICU at 32 and 36 weeks' postnatal menstrual age. They collected language and cognitive scores when the infants were 7 and 8 months of age. The findings showed that AWC in the NICU was positively correlated with later language and cognitive scores. In a recent large-scale longitudinal study conducted by the LENA Research Foundation, Gilkerson et al. (2018) collected monthly daylong recordings from 146 children (2 to 17 months, 18 to 24 months, and ≥ 25 months). Language and cognitive assessments were taken when the children were 9 to 14 years of age. They found that AWC and CTC at 18 to 24 months of age accounted for 3% to 30% and 23% to 37% of the variance in language and cognitive outcomes, respectively. However, no significant correlations were found for the 2 to 17 months and ≥ 25 months age groups.

Taken together, this body of research, in general, suggests specific associations between LENA's automated measures and child language and cognitive outcomes, providing further evidence to support Hart and Risley's findings that the early language environment plays a critical role in child language development. However, with regard to some LENA's automated measures, results have been inconsistent across studies. For example, whereas Gilkerson et al. (2018) found a significant correlation between AWC and language outcome measures, Ambrose et al. (2014) did not show such an association. Moreover, whereas Gilkerson et al. (2018) only found significant associations between AWC and outcome measures during the relatively narrow developmental window of 18 to 24 months of age, Caskey et al. (2014) and others (Adams et al., 2018; Greenwood et al., 2011) found such associations during much earlier period. These contrasting findings may be due to differences in methodologies, experimental design, population characteristics, among others, which render the interpretations of the results perplexing. Therefore, quantitatively combining effect sizes across studies using systematic review and meta-analysis methods will provide more accurate population estimates of the associations between LENA's automated measures and child language outcomes. In addition, given a great deal of variability with respect to methodologies and sample characteristics, it is relevant to test the moderation effects of these variables with respect to these associations between LENA's automated measures and language outcomes.

Purpose of the current study

The primary objective of this systematic review and meta-analysis was to evaluate the predictability of LENA's automated measures for language development in a variety of populations. Specifically, we asked to what extent do AWC, CTC, and CVC predict child language development? We predicted that these measures, in general, would significantly predict child language, although it is possible that some of these measures might show stronger predictive power. We also conducted moderator analyses to examine potential factors that might influence this relationship. Findings from this meta-analysis will have significant theoretical, methodological, and clinical significance. From a theoretical perspective, examining the specific measures of the early language environment that relate to child language would enhance our understanding of the developmental trajectory and provide insights into theories of child language acquisition. From a methodological perspective, although the primary goal of the current research was not to assess LENA's accuracy, the analysis of the predictability of LENA's automated measures for child language outcomes would serve as an indirect way to evaluate LENA's validity for estimating aspects of child language input, guiding the future exploration of language environment using LENA. The rationale was that if LENA algorithms provide reliable estimates of child language input, there would be a reliable predictive power of LENA's automated measures for child developmental outcomes. From a clinical perspective, current early intervention programs have recognized the importance of family-centered early intervention for children who are at risk for language delays (Moeller, Carr, Seaver, Stredler-Brown, & Holzinger, 2013). Therefore, knowledge about which measures are the most predictive of child language development would improve clinicians' ability to coach and encourage families to provide a language environment tailored to promote their child's language development.

Methods

Study database development

We followed the Preferred Reporting Items for Systematic Review and Meta-Analyses survey protocol (PRISMA; Moher, Liberati, Tetzlaff, & Altman, 2009) to conduct the analysis. Multiple methods were used to search for relevant literature. First, we put together a list of 20 reports based on the authors' knowledge of the literature on LENA or recommendations from colleagues. Second, we conducted exhaustive searches in English from 2008 to 2020 when LENA became commercial available. We used various combinations of terms including Language ENvironment Analysis, LENA, language environment, Adult Word Count, AWC, Conversational Turn, CTC, Child Vocalization, CVC, language development, language outcomes, and vocabulary on Science Direct, Pubmed, and scholar.google.com. Similar keywords were also used to set up Google scholar alerts for sending new results matching the search. The combination of our search methods allowed us to avoid potential bias in reporting (Rosenthal, 1979; Rosenthal & DiMatteo, 2001), as both published and unpublished research was included in search results from scholar.goole.com and Google scholar alerts. Unpublished research included dissertations and conference papers. We included 140 records in our database from these search methods. Third, we used the ancestral method (examining the reference section of relevant literature) and inspected publications listed on the LENA Research Foundation website; we identified an additional 11 records from this method. Together these search strategies resulted in a database of 171 records for further review. Among these compiled records, 8 duplicates were identified and were thus excluded. Figure 1 presents the PRISMA flow diagram.

Study selection criteria

Inclusion and exclusion criteria were determined based on study design, LENA measurements, child characteristics and age, and outcome measures.

Inclusion Criteria

(1) Study design: Studies must report a relation between at least one LENA's automated measure (AWC, CTC, CVC) and one language measure. LENA and language measures could be taken either concurrently or longitudinally. To maximize sample size, both vocabulary measures and general language measures were included. Moreover, language measures can be either receptive or expressive, and can be elicited from the children or reported by parents. See Table 1 for specific outcome measures for each study. (2) *Population.* We included children with a variety of developmental status, for example, typically-developing (TD), preterm, autism spectrum disorder (ASD), and hearing loss (HL). However, for children with HL, the LENA recording had to be taken after children received hearing devices. To be included, the mean or median age of child participants must be no greater than 48 months at the time of LENA measure, as the LENA system has not been validated on children older than 48 months.

Exclusion criteria—Intervention and qualitative studies were excluded. However, intervention studies were included if an association between LENA's automated measure and language was reported prior to intervention. To avoid risk for correlated measurement errors, we exclude the reports if their language measures were derived from the LENA. This was because when the language measures also derive from the same sources, the correlation between these two variables will be artificially elevated (Yoder & Symons, 2010). Single case studies and studies with a very small sample size (N < 5) were also excluded. This was because correlations from groups with a small number of participants, albeit strong, are difficult to assess if the correlation assumptions were met. We also excluded review articles and other non-primary reports (sample overlapped substantially with other records). Studies reporting statistics in other languages were not screened. Studies that only included manually coded measures from LENA recordings were excluded.

Study selection—During the first selection phase (Abstract phase), the first author screened the report abstracts following the inclusion and exclusion criteria. The second author randomly selected and independently screened 25% of all the abstracts in the database following the same criteria. There was 94.1% agreement on the inclusion of these reports during the Abstract phase. The disagreements were resolved by discussion and reviewing of the reports. During this first selection phase, the number of reports was reduced from 163 to 59. For a study to be excluded during this stage, the title and the abstract had to clearly indicate that the study failed to meet at least one of the inclusion criteria. When the title or the abstract did not clearly indicate whether it met the criteria, reports were retrieved for full review during the second phase (Full paper phase). During the Full paper phase, the full-text reports were retrieved for all the 59 records that passed the Abstract phase. The number of records was reduced from 59 to 17 for use in the meta-analysis. Studies removed following application of each criterion according to PRISMA guidelines are summarized in Figure 1.

Study coding and data extraction

Data were extracted from the results reported in each study to allow for the calculation of the effect size and its variance. We created and used a data-extraction form to organize bibliographic and study characteristics, including participants, methodological variables, and effect sizes. Participant variables included mean/median age of participants at LENA assessment, developing status (TD, ASD, preterm, HL, mixed, etc), gender, SES, primary language, age at language outcome measures. Four studies included more than one population but collapsed the population in the analyses; these studies were coded as mixed in the moderator analysis. Methodological variables included sample size, automated LENA measures examined, AWC/hour, CTC/hour, CVC/hour, number of LENA recordings, average duration of LENA recordings, duration between LENA recordings and language outcomes, and measures for language. For language measures, we coded both modality and method of language assessment. This decision was motivated by findings that expressive and receptive language skills may be related to different constructs and rely on different underlying representations (Chang, Dell, & Bock, 2006), as well as that parent report and direct child assessment may have different validity in evaluating child language abilities (Sachse & Von Suchodoletz, 2008). Language measure was coded as expressive or receptive; when a study employed both expressive and receptive language measures, this study was coded as mixed; moreover, language measure was coded as parent report or direct child assessment; if a study employed both parent report and direct child assessment, the study was coded as mixed. The complete list of reports included in the meta-analysis and the variables coded is shown in Table 2.

Analytical strategies

Effect size calculation—Effect sizes were computed on the basis of Pearson's *r*, partial *r*, or other convertible statistics, such as *p*s and beta, when the correlation coefficients were not available (Peterson & Brown, 2005; Rosenthal & DiMatteo, 2001). When the statistics were not reported in the original reports, the first author contacted the corresponding authors of these reports for more details. Positive effect sizes indicate that the association was in the predicted direction, whereas negative effect sizes indicate that the association was in the opposite direction to the hypothesis. The interpretation of effects sizes were based on Cohen (1988)'s conventions: small (Pearson's $r = .1$), medium (Pearson's $r = .3$), or large (Pearson's $r = .5$).

Only one effect size per independent participant group per measure was considered in subsequent analyses. Under the circumstances when a given study reported multiple correlation coefficients for a particular measure from a single participant group, the weighted average age at LENA recording, the weighted average age at language measures, and a weighted mean *r* was calculated, where the weight was based on the sample size in each correlation (Colonnesi, Stams, Koster, & Noom, 2010; Cristia, Seidl, Junge, Soderstrom, & Hagoort, 2014; Milligan, Astington, & Dack, 2007; Rosenthal & DiMatteo, 2001). A couple of reports included multiple effect sizes on different samples. Consequently, there was not a direct correspondence between the number of reports and the number of effect sizes. Our meta-analysis included 17 reports with 18 samples and 40 unique effect sizes, 17 for AWC, 13 for CTC, and 10 for CVC.

Following current meta-analytic standards, all correlation coefficients were converted to the Fisher's *z* scale to account for each effective size by using the formula: $z = 0.5 \times ln\left(\frac{1 + r}{1 - r}\right)$.

(Borenstein, Hedges, Higgins, & Rothstein, 2011). All analyses were performed using the z-transformed values. Effect sizes were transformed back to *r* for reporting.

Robust variance estimation—Because most of the individual reports included more than one LENA's automated measure, the effect sizes for the current analysis were not independent of each other. Therefore, we used robust variance estimation to address the dependent effect sizes issue (Hedges, Tipton, & Johnson, 2010). Accordingly, we used random effects models with approximately inverse variance weights. Under the random effects model, the true effects are assumed to vary between studies and the summary effect is the weighted average of the effects across different studies (Borenstein et al., 2011).

Heterogeneity analyses—We conducted chi-square tests of the Q and their ps to evaluate the statistical significance of heterogeneity and the magnitude of heterogeneity using the I^2 value (Borenstein et al., 2011). Following Higgins, Thompson, Deeks, and Altman (2003)'s guideline, I^2 is interpreted as low (.25), moderate (.50), or high (.75).

Sensitivity analyses—We conducted sensitivity analyses by assessing the impact of the individual effect size on the main effects by removing one effect size and while keeping the rest of other effect sizes constant at a point from the analysis. If results remain consistent across the different analyses, then the results can be considered robust and representative.

Publication bias—Publication bias is a major risk to the validity of meta-analysis. It occurs when the outcome of a study influences the decision of whether or not to publish it, and this is especially true for studies with a small sample size. Due to publication bias, statistically significant results are more likely to be published than those with null results (Rosenthal, 1979). We assessed risk for publication bias by using funnel plots and the Egger's test. Funnel plots provide a visual method for evaluating whether selection/publication bias exists. A funnel plot shows effect sizes versus a metric of standard error. It was calculated by using the formula: $se_r = \sqrt{\frac{1 - r^2}{n - 2}}$, where r is the correlation coefficient, and n is the sample size. If there is a publication bias in the literature, we should expect studies to be asymmetrically distributed around the weighted average effect size, with more variability for less precise studies. Thus, a study with relatively low precision will have a larger standard error than a study with relatively high precision. We used Egger's tests for funnel plot asymmetry. Non-significant asymmetry indicated that selection/publication bias was not found in our systematic search.

Results

Study Characteristics

Table 1 presents the characteristics of all included studies. The studies took place in 5 countries and included speakers of 3 primary languages. The majority of the studies were conducted in the US (13 out of 17), and most studies included English-speaking participants (15 out of 17). Sample sizes ranged from 8 to 306; participant ages ranged from −1.59 months (adjusted age for preterm infants) to 46.93 months. A total of 1,093 participants were included in the final analysis. The latency between the LENA recording and language assessment ranged from concurrent to 140.88 months.

Meta-analytic analysis results

We conducted a random effects model meta-analysis grouped by measures (AWC, CT, CV) in the R environment (R Core Team, 2014). To estimate whether the overall correlation was statistically significant at the level of zero, effect sizes were pooled across reports to obtain a single meta-analytic estimate. This meta-analytic effect size can be thought of as the best estimate of the effect size for a phenomenon under examination.

The mean overall effect size and effect sizes for each measure are presented in Figure 2. The overall mean effect size for the combined correlation between LENA's automated measures and language outcomes was significant, $r = 0.27$, 95% CI [0.22, 0.31], $z = 10.77$, $p < .001$. This correlation was considered medium based on Cohen's convention. Heterogeneity was small, $I^2 = .12$, $Q(39) = 44.26$, $p = .259$, suggesting that there was not significant variability in effect sizes between studies. To determine whether an individual effect size would significantly influence the overall mean effect size, for each measure, we ran a sensitivity analysis in which each effect size was excluded one at a time using metaif() function (Schwarzer & Schwarzer, 2012). Results indicated that the finding was robust to the removal of any individual effect size, $0.26 < r < 0.28$, ps < .0001. suggesting that no single sample significantly influenced the overall effect size. Egger's tests for funnel plot asymmetry were nonsignificant, $z = .97$, $p = .330$. Non-significant asymmetry indicated that selection/ publication bias was not found in our systematic search. Figure 3 presents the funnel plot.

We conducted further analyses to assess whether one of the three LENA's automated measures differed in their predictive power for language outcomes. A test for subgroup differences showed a trend for differences among the three measures, $Q(2) = 4.94$, $p = .085$, suggesting at least one of the three measures might be better or poorer in predicting language outcomes. Therefore, we further examined the mean effect sizes for AWC, CTC, and CVC, separately. Figure 2 presents meta-analytic effect size estimates for each of the three measures. The mean effect size for the correlation between AWC and language outcomes was $r = 0.21$, 95% CI [0.14, .27], $z = 6.11$, $p < .0001$, indicating that AWC was significantly correlated with language outcomes with a small-to-medium effect size. The overall mean effect size for the correlation between CTC and language outcomes was $r = 0.31$, 95% CI [0.21, 0.40], $z = 5.98$, $p < .0001$, suggesting that CTC was significantly correlated with language outcomes with a medium effect size. The overall mean effect size for the correlation between CVC and language outcomes was $r = .32$, 95% CI [0.21, 0.42], $z = 5.34$, $p < .0001$, suggesting that CVC was significantly correlated with language outcomes with a medium effect size.

Moderator analyses

An additional set of analyses was conducted to test potential factors affecting the relationship between LENA's automated measures and child language outcomes using meta-regression for continuous variables and subgroup analyses for categorical variables. According to Bakermans-Kranenburg, Van Ijzendoorn, and Juffer (2003)'s guideline, for categorical factors, moderator analysis is appropriate only when at least two of the subsets consisted of at least four effect sizes. For each variable, we excluded from the analysis those effect sizes where the dependent measure included a mixed population group, or language measure that combined receptive and expressive language, or parent report and direct child assessment. Therefore, we conducted 6 moderator analyses that fulfilled this criterion: 2 continuous moderators including age at LENA recording and the gap between LENA recording and language outcomes; 4 categorical moderators including child developmental status (TD, preterm, ASD, HL), publication status (published, unpublished), language test modality (receptive, expressive), and language test method (parent report, direct child assessment).

The results of the moderator analyses are reported in Table 2. None of the four categorical moderators, child developmental status, $Q(3) = 1.17$, $p = .759$, publication status, $Q(1) = .28$, $p = .595$, language test modality, $Q(1) = 2.56$, $p = .110$, or language test method $Q(1) = .43$, $p = .511$, were significant moderators. Moreover, age at recording was not significant, $Q(1) = .00$, $p = .984$. These findings suggest that overall the associations between the LENA's automated measures and child language skills are not necessarily influenced by these factors. However, the gap between the LENA recording and language assessment, although not significant, seemed to influence the association between LENA's automated measures and child language measures, $Q(1) = 2.66$, $p = .103$. Specifically, this association weakened with an increase in gap, as shown in Figure 4. It should be noted that the gap between the LENA recording and language measures was exceptionally long in Gilkerson et al. (2015)'s report as compared to other included reports, which might be a potential outlier. Therefore, we ran an additional moderator analysis excluding Gilkerson et al. (2015)'s report. The results showed that the gap was a significant moderator, $Q(1) = 6.96$, $p = .008$. Taken together, these findings suggest that the association between LENA's automated measures and child language measures decreases with an increase in the gap.

Discussion

Findings of the Meta-analysis

Early language environment plays a critical role in child language development. The advent of the LENA system allows for an automated measurement of various aspects of the child language environment. The purpose of the current study was to use meta-analytic techniques to assess the predictive strength of the LENA automated indices of the early language environment, specifically, AWC, CTC, and CVC, on child language. Overall, we found a medium significant association between LENA's automated measures and child language outcomes across the three measures. These findings lend additional support for the relationship between early language environment and language outcomes. We can be confident in the attested relationship as our analyses showed a low magnitude of heterogeneity and withstood the sensitivity test. Moreover, the findings did not provide any evidence of publication bias, as suggested by the funnel plot and asymmetry test. Among the three automated measures, CTC and CVC showed medium size significant associations with language, whereas AWC showed a small-to-medium size significant association with language. These findings are in general consistent with and supported by previous theoretical and empirical evidence, which we discuss below.

The positive relationship between conversational turns and language skills suggests that everyday interactions between caregivers and their children may be particularly important for language development. According to theories of social learning, language learning relies heavily on children's sensitivity to joint attention and their participation in the social communication (Baldwin, 1995; Morales et al., 2000; Salo, Rowe, & Reeb-Sutherland, 2018). On the one hand, conversational turns provide increased opportunities and multimodal cues for children to exploit in the service of language learning; the increased language exposure along with children's elevated attention during the interactions with caregivers, lead to an overall increase in the quantity of quality of information processed. On the other hand, by engaging in responsive and reciprocal modes of interactions, children convey their preferences of communication to caregivers, who adjust and provide contingent real-time linguistic and pragmatic cues that could facilitate language learning. A considerable body of research during the past 30 years has provided empirical supporting evidence. For example, Tamis-LeMonda, Bornstein, and Baumwell (2001) showed that maternal responsiveness at 9 and 13 months significantly predicted the timing of language milestones.

In addition, neural evidence suggests that high-quality communications provide a positive social feedback loop supporting the development of brain areas involved in speech and language learning (Kuhl & Rivera-Gaxiola, 2008; Romeo et al., 2018). In a recent study, Romeo et al. (2018) examined the relationship between child natural language experience, neural responses during language processing, and linguistic skills. They demonstrated that children who experienced more conversational turns at home exhibited greater activation in Broca's area during language processing, which mediated the relationship between children's language exposure and language abilities.

Moreover, due to the nature of CTC, it is likely to include a high proportion of child-directed speech, a speech style has been shown to benefit child speech processing and language development (Cristia & Seidl, 2014; Drotar & Sturm, 1988; Song, Demuth, & Morgan, 2010; Trainor, Austin, & Desjardins, 2000). Child-directed speech is characterized by slower speaking rate, higher pitch, wider pitch range, longer pauses, and expanded vowel space (Burnham, Kitamura, & Vollmer-Conna, 2002; Cristia, 2010; Fernald & Simon, 1984; Papoušek & Hwang, 1991; Wang, Lee, & Houston, 2016). These unique properties of child-directed speech are shown to engage and sustain attention, allowing infants and children more opportunities to access, encode, and process speech. Therefore, the child-directed speech that accompanies caregiver-child interactions provides additional support for language learning. Taken together, a larger number of CTC reflects a higher degree of social engagement between caregivers and children, which benefits child language development through multiple mechanisms.

The significant correlation between child vocalization and other measures of child language is also in line with previous theoretical and empirical evidence (J. McDaniel, Slaboch, & Yoder, 2018; Moeller et al., 2007). Prior to producing meaningful utterances, infants gain fundamental gross motor skills and progress through a continuum of predictable and universal prelinguistic vocal stages, beginning with non-speech-like vocalizations and transitioning to more complex and speech-like vocalizations (Nathani, Ertmer, & Stark, 2006; Oller, 2000; Stoel-Gammon, 2011). Development of child vocalization is not an isolated phenomenon, but rather, is related to production patterns of early words and later spoken language development (Kent & Miolo, 2017; Menyuk, Liebergott, & Schultz, 2014; Oller, 1978; Vihman, Macken, Miller, Simmons, & Miller, 1985). For example, in a longitudinal study including 53 children, Menyuk et al. (2014) examined the children's vocal development over the first 3 years of life. They showed that the rate at which the children shifted from vocalization to babbling was related to the rate at which they achieved the mastery of articulating consonantal sounds, which in turn, was related to the rate of word acquisition and morpheme development.

Recent research exploring the mechanisms underlying vocal development has proposed that child vocal development is, at least in part, driven by social interaction. According to transactional hypotheses (Harding, 1983; Moeller et al., 2007; PapouSek, 1993), children learn language through bidirectional and transactional exchanges with caregivers. Children use vocalizations to respond to and participate in the interaction with adults, and child vocalizations often reflect parent word models. This dynamic reciprocal relationship also means that child vocalizations could elicit developmentally appropriate responses from caregivers, which in turn leads to greater child vocal skills that subsequently receive more complex language input from the caregiver. For example, infants who received contingent responses on their babblings showed a rapid reconstruction of their vocalizations, incorporating the phonological patterns from the caregivers' responses (Goldstein & Schwade, 2008). Moreover, infants tend to produce more frequent speech-like vocalizations if they receive contingent responses from their caregivers than if they receive non-contingent responses (Goldstein et al., 2003).

Finally, the small-to-medium size association between AWC and language outcomes suggest that AWC has less predictive power for language outcomes compared to the CTC and CVC measures. These findings are not surprising as AWC includes both speech directed to children and overheard by children, and the relationship between the amount of speech to children and overheard by children is shown to be complementary in nature and changes as a function of child development (Bergelson et al., 2019). While the quantity of input is clearly important, and children can learn from both child-directed speech and overheard speech, much recent evidence suggests that child-directed speech appears to play a more important role in language development (Golinkoff, Hoff, Rowe, Tamis-LeMonda, & Hirsh-Pasek, 2018; Weisleder & Fernald, 2013). Supporting this, empirical findings suggest that infants who experienced a larger amount of child-directed speech at home, but not a larger amount of overheard speech, became more efficient in lexical processing and had a larger expressive vocabulary by 24 months of age (Weisleder & Fernald, 2013).

Notably, the moderator analyses showed that the relationship between LENA's automated measures and language skills was robust and largely consistent regardless of child developmental status (TD, preterm, ASD, HL), publication status (published, unpublished), language test modality (receptive, expressive), language test method (parent report, direct child assessment), or the age at which the LENA recordings were collected. However, we observed a moderation effect of the gap between the time when LENA recordings were taken and the time when language was measured, reflecting a decrease in effect sizes as the gap increased. These findings may suggest that LENA's automated measure have a less predictive strength on long-term language outcomes.

Theoretical, methodological, and clinical implications

These results have significant theoretical, methodological, and clinical implications. First, from a theoretical perspective, these findings support Hart and Risley's and other's findings that early language environment has a significant impact on child language development (e.g., Gilkerson et al., 2018; Hart & Risley, 1995; Weisleder & Fernald, 2013). These findings also extend findings from previous work which only included small samples and short recordings. Because LENA collects massive speech samples from home, it allows for a generalization of the previous findings to the naturalistic language environment.

Second, from a methodological perspective, these findings provide indirect evidence for the predictive validity of the LENA system to automatically analyze language environments with sufficient accuracy to detect individual differences that correlate with language skills. Moreover, the findings provide valuable information for child language researchers who seek to identify the best practices into integrating the LENA system into the exploration of child language acquisition.

From a clinical perspective, these results support using LENA as a potential tool for clinicians working with young children and their families. Although Hart and Risley (1995) showed that the early language environment plays a critical role in child development, the methodological limitations related to laborious transcriptions severely limited its clinical application. The LENA system provides an alternative to manually analyze the naturalistic language environment. The findings that LENA's automated measures significantly predicted child language outcomes have profound implications for early intervention programs for identifying children who might be at risk for poor language development and providing appropriate services to improve the home language environment of young children (Suskind et al., 2016; Suskind et al., 2013).

Limitations of the LENA system

Despite the advantages and massive potentials that the LENA system offers to researchers and clinicians, it has several limitations that would benefit from future development. First of all, the LENA System is only normed for children up to 48 months of age. Although this does not necessarily suggest that the LENA system is not valid for children over 48 months of age, further evaluation is encouraged to assess whether the LENA's reliability or accuracy for older children. This is important as it will allow for an examination of the change and/or stability of children's auditory environment across development and explore how these features may be related to the growth of child language skills.

Second, LENA also makes labeling errors which changes as a function of talker gender and speech register (Bulgarelli & Bergelson, 2019; Gilkerson et al., 2015; Lehet, Arjmandi, Dilley, & Houston, under review; VanDam & Silbert, 2016; Xu et al., 2009). This could affect the AWC, CTC, and CVC estimates, as these calculations might a priori depend on LENA's classification accuracy of segments (Kimbrough Oller, 2010). For example, approximately one-third of child speech labeled by LENA were annotated as adult speech by human coders, especially when the female adult speakers raised their voices when talking to children, resulting in reduced AWC estimates in the output. In addition, the LENA system does not identify speakers when speech and sounds overlap (e.g., speech + speech, speech + noise, and noise + noise); instead, all these combinations are categorized as overlapping speech (OLN), which is not included in the calculation of AWC (Gilkerson et al., 2008). This could reduce LENA' accuracy in analyzing AWC for the recordings taken from a busy household with a many family members, or classrooms with elevated noise levels and multiple talkers speaking at the same time. Therefore, improvement in current speech processing technology for extracting information from audio recordings with higher accuracy would greatly benefit and advance research using rich naturalistic language data.

Moreover, some research questions require more fine-grained annotations than the current LENA system can provide. For example, the current LENA system does not distinguish between the speech directed to children and the speech overhead by children; this distinction is critical given the significant role of IDS on language development (Weisleder & Fernald, 2013). Moreover, the LENA system does not distinguish between canonical babbling (which involves well-formed syllables) and other types of speech-like vocalizations. Canonical babbling has been found to be significantly correlated with later speech and language development and to predict developmental disorders (Chapman, Hardin-Jones, & Halter, 2003; Oller, Eilers, Steffens, Lynch, & Urbano, 1994). Another aspect of limitation is that the LENA system does not provide speech transcription, which is essential but probably the most time-consuming component, for those researchers who examine the vocabulary, grammatical, and syntactic structures of speech. Therefore, one important future development of the LENA system would be to provide more fine-grained annotations and transcriptions to suit broader needs.

These limitations require joint efforts between speech technology and research communities to develop more accurate and comprehensive systems in the future. Nevertheless, in the meantime, researchers who are in need of the information that LENA does not currently provide LENA may still take advantage of the LENA output and adopt integrated methods to obtain desired measures. For example, to calculate amount of infant-directed speech and overheard speech from LENA recordings, Weisleder and Fernald (2013) coded each 5-min segment produced by the LENA system as either predominantly infant-directed speech or overhead speech; based on the AWC generated by LENA for each segment, they obtained estimates of amount of infant-directed speech and overhead speech. To examine North American children's early auditory environment, Bergelson et al. (2019) selected 20 LENA conversational blocks each of which contained at least 10 FAN (Female-Adult Near) or MAN (Male-Adult-Near) segments as annotated by LENA. They then manually tagged each segment for speaker gender (Female vs. Male) and addressee (Adult vs. Child) which allowed for a comparison of speech produced by females vs. adults, and speech directed to children vs. overhead speech. These integrated methods that leverage the LENA output and incorporate a reasonable amount of human annotations, albeit may not exact or perfect, improve the accuracy of estimates and broaden the prospect of using LENA for various research purposes.

Limitations of the current meta-analysis

Despite the contributions of this meta-analysis to the field, we acknowledge several limitations of this work. First of all, as we mentioned earlier that 13 out of the 15 reports were conducted with English-speaking children. Due to the small number of studies with non-English-speaking children, we were unable to examine whether language background would moderate the predictability of LENA's automated measures for language. Therefore, it is still inconclusive whether the LENA system would have similar predictive strength for language outcomes in children who speak languages other than English. Expanding the use of the LENA System with children from different language background is necessary to assess the reliability and accuracy of the LENA system across cultures.

Second, although we made an effort to exclude secondary-reports whose samples substantially overlapped with other included reports, the samples were not always precisely described in the reports; consequently, it is possible that some children were included in the association calculation more than once. This is a general issue in meta-analytical work, which would benefit from detailed descriptions and documentation of participant characteristics. Third, it is possible that in addition to the 5 factors we have examined in this analysis, other factors may also moderate the relationship between LENA's automated measures and language outcomes. For example, previous research has demonstrated that the quantity of parental talk and interaction in the early language environment are correlated with SES (Hart & Risley, 1995); however, we were unable to conduct moderator analysis on SES because it was either not reported, or assessed by different criteria. Therefore, future studies should examine other potential factors that may moderate this relationship. Finally, due to the correlational nature of the findings, the direction of the relationship cannot be established. Future intervention studies with randomized control trial designs are encouraged to elucidate the potential causal effects of early environmental factors on child language development.

Future directions

In addition to the automatic processing of linguistic input, LENA also provides estimates of other auditory information in a child's language environment, including overlapping noise, TV and media, and other noises. While past research in the past has almost exclusively focused on the aspects of early language input that benefit child language development, recent research has demonstrated that specific factors from the early auditory environment may have a deleterious impact on child speech processing and language outcomes. For example, Ambrose et al. (2014) showed that toddlers with hearing loss who were exposed to a larger amount of electronic media showed poorer receptive language skills. Similarly, Williams, Wang, Dilley, and Houston (2019) showed that the total amount of auditory chaos, defined by the total amount of overlapping noise, TV or media, and other noise calculated by LENA, was negatively correlated with child speech processing efficiency. Therefore, one of the important future directions is to examine the relationship between positive and negative aspects of child home language environment, and how these factors may interact to explain variability in child language development.

Finally, using LENA as a tool for intervention is another important future direction. Given the important role of early language environment on child language, cognitive and social development, early intervention programs focusing on teaching and educating parents to improve their speech input to and interaction with their children is crucial. The LENA system offers a potential source of feedback on the magnitude of change in parents' behavior. Recently, researchers have begun to explore this possibility and showed significant elevations in parent talk and interaction with their children using LENA feedback along with parent coaching in home visiting programs (Suskind et al., 2016; Suskind et al., 2013) and online intervention programs (Gilkerson, Richards, & Topping, 2017). Future studies including diverse populations and language to investigate how parent behavioral change might be related to child language development would provide important knowledge to inform evidence-based early intervention.

Conclusions

This systematic review and meta-analysis provides the first qualitative assessment of the predictability of LENA's automated measures, AWC, CTC, and CVC, for child language outcomes. The findings extend previous LENA validity literature by documenting a moderate association between LENA's automated measures and language outcomes. These findings confirm the predictive validity of LENA's automated measures for child language development. Moreover, these findings will inform early intervention strategies that use LENA as a tool to engage and measure parental language input to children. Although we are optimistic about the use of the LENA system for both research and clinical purposes, we have highlighted specific areas that require future development and research areas that would benefit from the use of the LENA system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute on Deafness and Other Communication Disorders Grant [R01 DC008581] to Derek M. Houston and Laura Dilley. The authors would like to thank Katherine Adams, Virginia Marchman, Jena McDaniel, and Eric Walle, for providing additional correlational coefficients and detailed descriptions of their data.

This research was supported in part by National Institute on Deafness and Other Communication Disorders Grant (R01 DC008581). We have no conflict of interest to declare.

Reference

Adams KA, Marchman VA, Loi EC, Ashland MD, Fernald A, & Feldman HM (2018). Caregiver talk and medical risk as predictors of language outcomes in full term and preterm toddlers. Child Development, 89(5), 1674–1690. [PubMed: 28452393]

Ambrose SE, VanDam M, & Moeller MP (2014). Linguistic input, electronic media, and communication outcomes of toddlers with hearing loss. Ear and Hearing, 35(2), 139–147. doi:10.1097/AUD.0b013e3182a76768 [PubMed: 24441740]

Bakermans-Kranenburg MJ, Van Ijzendoorn MH, & Juffer F (2003). Less is more: meta-analyses of sensitivity and attachment interventions in early childhood. Psychological Bulletin, 129(2), 195–215. [PubMed: 12696839]

Baldwin DA (1995). Understanding the link between joint attention and language Joint attention: Its origins and role in development. (pp. 131–158). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Bergelson E, Casillas M, Soderstrom M, Seidl A, Warlaumont AS, & Amatuni A (2019). What do North American babies hear? A large-scale cross-corpus analysis. Developmental Science, 22(1), e12724. [PubMed: 30369005]

Borenstein M, Hedges LV, Higgins JP, & Rothstein HR (2011). Introduction to meta-analysis. Chichester: John Wiley & Sons.

Brainard MS, & Doupe AJ (2000). Auditory feedback in learning and maintenance of vocal behaviour. Nature Reviews Neuroscience, 1(1), 31–41. [PubMed: 11252766]

Bulgarelli F, & Bergelson E (2019). Look who's talking: A comparison of automated and human-generated speaker tags in naturalistic day-long recordings. Behavior Research Methods, 1–13. [PubMed: 29967978]

Burnham D, Kitamura C, & Vollmer-Conna U (2002). What's New, Pussycat? On Talking to Babies and Animals. Science, 296(5572), 1435. [PubMed: 12029126]

Busch T, Sangen A, Vanpoucke F, & van Wieringen A (2018). Correlation and agreement between Language ENvironment Analysis (LENA™) and manual transcription for Dutch natural language recordings. Behavior Research Methods, 50(5), 1921–1932. [PubMed: 28936690]

Caskey M, Stephens B, Tucker R, & Vohr B (2014). Adult talk in the NICU with preterm infants and developmental outcomes. Pediatrics, 133(3), e578–e584. [PubMed: 24515512]

Chang F, Dell GS, & Bock K (2006). Becoming syntactic. Psychological review, 113(2), 234. [PubMed: 16637761]

Chapman KL, Hardin-Jones M, & Halter KA (2003). The relationship between early speech and later speech and language performance for children with cleft lip and palate. Clinical Linguistics & Phonetics, 17(3), 173–197. [PubMed: 12858838]

Christakis DA, Gilkerson J, Richards JA, Zimmerman FJ, Garrison MM, Xu D, … Yapanel U (2009). Audible television and decreased adult words, infant vocalizations, and conversational turns: a population-based study. Archives of Pediatrics & Adolescent Medicine, 163(6), 554–558. [PubMed: 19487612]

Cohen J (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum Associates.

Colonnesi C, Stams GJJ, Koster I, & Noom MJ (2010). The relation between pointing and language development: A meta-analysis. Developmental Review, 30(4), 352–366.

Cristia A (2010). Phonetic enhancement of sibilants in infant-directed speech. The Journal of the Acoustical Society of America, 128(1), 424–434. [PubMed: 20649236]

Cristia A, Bulgarelli F, & Bergelson E (2020). Accuracy of the Language Environment Analysis System Segmentation and Metrics: A Systematic Review. Journal of Speech, Language, and Hearing Research, 1–13.

Cristia A, & Seidl A (2014). The hyperarticulation hypothesis of infant-directed speech. Journal of Child Language, 41(04), 913–934. [PubMed: 23406830]

Cristia A, Seidl A, Junge C, Soderstrom M, & Hagoort P (2014). Predicting individual variation in language from infant speech perception measures. Child Development, 85(4), 1330–1345. [PubMed: 24320112]

Drotar D, & Sturm L (1988). Prediction of intellectual development in young children with early histories of nonorganic failure-to-thrive. Journal of Pediatric Psychology, 13(2), 281–296. [PubMed: 3171820]

Elo H (2016). Acquiring Language as a Twin: Twin children's early health, social environment and emerging language skills. (Unpublished doctoral dissertation), University of Tampere.

Fernald A, & Simon T (1984). Expanded intonation contours in mothers' speech to newborns. Developmental Psychology, 20(1), 104–113. doi:10.1037/0012-1649.20.1.104

Ganek H, & Eriks-Brophy A (2018). Language ENvironment analysis (LENA) system investigation of day long recordings in children: A literature review. Journal of Communication Disorders, 72, 77–85. [PubMed: 29402382]

Gilkerson J, Coulter K, & Richards J (2008). Transcriptional analyses of the LENA Foundation natural language corpus (LENA Foundation Technical Report LTR-06–2). Retrieved from LENA Foundation: http://www/.lenafoundation.org/TechReport.aspx/Transcription/LTR-06-2.

Gilkerson J, Richards JA, & Topping K (2017). Evaluation of a LENA-based online intervention for parents of young children. Journal of Early Intervention, 39(4), 281–298.

Gilkerson J, Richards JA, Warren SF, Montgomery JK, Greenwood CR, Kimbrough Oller D, … Paul TD (2017). Mapping the early language environment using all-day recordings and automated analysis. American Journal of Speech-language Pathology, 26(2), 248–265. [PubMed: 28418456]

Gilkerson J, Richards JA, Warren SF, Oller DK, Russo R, & Vohr B (2018). Language experience in the second year of life and language outcomes in late childhood. Pediatrics, 142(4), e20174276. [PubMed: 30201624]

Gilkerson J, Zhang Y, Xu D, Richards JA, Xu X, Jiang F, … Topping K (2015). Evaluating language environment analysis system performance for Chinese: A pilot study in Shanghai. Journal of Speech, Language, and Hearing Research, 58(2), 445–452.

Goldstein MH, King AP, & West MJ (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. Proceedings of the National Academy of Sciences, 100(13), 8030–8035.

Goldstein MH, & Schwade JA (2008). Social feedback to infants' babbling facilitates rapid phonological learning. Psychological Science, 19(5), 515–523. [PubMed: 18466414]

Golinkoff RM, Hoff E, Rowe ML, Tamis-LeMonda CS, & Hirsh-Pasek K (2018). Language Matters: Denying the Existence of the 30-Million-Word Gap Has Serious Consequences. Child Development, Advanced online print.

Greenwood CR, Schnitz AG, Irvin D, Tsai SF, & Carta JJ (2018). Automated language environment analysis: A research synthesis. American Journal of Speech-language Pathology, 27(2), 853–867. [PubMed: 29594313]

Greenwood CR, Thiemann-Bourque K, Walker D, Buzhardt J, & Gilkerson J (2011). Assessing children's home language environments using automatic speech recognition technology. Communication Disorders Quarterly, 32(2), 83–92.

Harding CG (1983). Setting the stage for language acquisition: Communication development in the first year In Golinkoff RM (Ed.), The transition from prelinguistic to linguistic communication (pp. 93–111). Hillsdale, NJ: Lawrence Erlbaum.

Hart B, & Risley TR (1995). Meaningful differences in the everyday experience of young American children. Baltimore, MD: Paul H Brookes Publishing.

Hedges LV, Tipton E, & Johnson MC (2010). Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods, 1(1), 39–65. [PubMed: 26056092]

Higgins JP, Thompson SG, Deeks JJ, & Altman DG (2003). Measuring inconsistency in meta-analyses. British Medical Association, 327(7414), 557–560.

Hoff E, & Naigles L (2002). How children use input to acquire a lexicon. Child Development, 73(2), 418–433. [PubMed: 11949900]

Huttenlocher J, Vasilyeva M, Cymerman E, & Levine S (2002). Language input and child syntax. Cognitive Psychology, 45(3), 337–374. [PubMed: 12480478]

Kent RD, & Miolo G (2017). Phonetic abilities in the first year of life. The handbook of child language, 303–334.

Kimbrough Oller D (2010). All-day recordings to investigate vocabulary development: A case study of a trilingual toddler. Communication Disorders Quarterly, 31(4), 213–222. [PubMed: 21037965]

Kuhl PK, & Rivera-Gaxiola M (2008). Neural substrates of language acquisition. Annu Rev Neurosci, 31, 511–534. doi:10.1146/annurev.neuro.30.051606.094321 [PubMed: 18558865]

Lehet M, Arjmandi MK, Dilley L, & Houston DM (under review). Accuracy of the Language ENvironment Analysis (LENA) system for quantifying adult speech experienced by infants in naturalistic settings.

McDaniel J, Slaboch KDA, & 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. [PubMed: 29195157]

McDaniel J, Yoder P, Estes A, & Rogers SJ (2019). Comparing the Construct Validity of Conventional and Automated Vocalization Measures for Young Children with Autism Spectrum Disorder. Paper presented at the 52nd Gatlinburg Conference on Research and Theory in Intellectual and Developmental Disabilities, Texas: San Antonio.

McDaniel R, & Purdy SC (2013). Using the Language Environment Analysis (LENA™) system to investigate the language environment and outcomes of deaf children: A pilot study. New Zealand Journal of Speech-Language Therapy, 67, 31–47.

Menyuk P, Liebergott JW, & Schultz MC (2014). Early language development in full-term and premature infants: Psychology Press.

Milligan K, Astington JW, & Dack LA (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. Child Development, 78(2), 622–646. [PubMed: 17381794]

Moeller MP, Carr G, Seaver L, Stredler-Brown A, & Holzinger D (2013). Best practices in family-centered early intervention for children who are deaf or hard of hearing: An international consensus statement. Journal of Deaf Studies and Deaf Education, 18(4), 429–445. [PubMed: 24030911]

Moeller MP, Hoover B, Putman C, Arbataitis K, Bohnenkamp G, Peterson B, … Stelmachowicz P (2007). Vocalizations of infants with hearing loss compared with infants with normal hearing: Part II–Transition to words. Ear and Hearing, 28(5), 628–642. [PubMed: 17804977]

Moher D, Liberati A, Tetzlaff J, & Altman DG (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of Internal Medicine, 151(4), 264–269. [PubMed: 19622511]

Morales M, Mundy P, Delgado CE, Yale M, Messinger D, Neal R, & Schwartz HK (2000). Responding to joint attention across the 6-through 24-month age period and early language acquisition. Journal of Applied Developmental Psychology, 21(3), 283–298.

Nathani S, Ertmer DJ, & Stark RE (2006). Assessing vocal development in infants and toddlers. Clinical Linguistics & Phonetics, 20(5), 351–369. [PubMed: 16728333]

Oller DK (1978). Infant vocalization and the development of speech. Allied Health and Behavioral Sciences, 1(4), 523–549.

Oller DK (2000). The emergence of the speech capacity: Psychology Press.

Oller DK, Eilers RE, Steffens ML, Lynch MP, & Urbano R (1994). Speech-like vocalizations in infancy: an evaluation of potential risk factors. Journal of Child Language, 21(1), 33–58. [PubMed: 8006094]

PapouSek M (1993). Stages of infant-directed speech: Relation to stages in the infant's interactional vocal repertoire. Paper presented at the Poster presented at the Society for Research in Child Development, New Orleans.

Papoušek M, & Hwang S-FC (1991). Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. Applied Psycholinguistics, 12(04), 481–504. doi:10.1017/S0142716400005889

Patterson SY (2010). An examination of parent-child dyadic interaction and the impact of a parent-training program designed to support the language development of toddlers diagnosed with autism spectrum disorder (ASD). (Unpublished Master's thesis), University of Alberta.

Peterson RA, & Brown SP (2005). On the use of beta coefficients in meta-analysis. Journal of Applied Psychology, 90(1), 175–181. [PubMed: 15641898]

Pretzer GM, Lopez LD, Walle EA, & Warlaumont AS (2019). Infant-adult vocal interaction dynamics depend on infant vocal type, child-directedness of adult speech, and timeframe. Infant Behavior and Development, 57, 101325. [PubMed: 31100586]

R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ramírez-Esparza N, García-Sierra A, & Kuhl PK (2017). The Impact of Early Social Interactions on Later Language Development in Spanish-English Bilingual Infants. Child Development, 88(4), 1216–1234. doi:10.1111/cdev.12648 [PubMed: 27759883]

Romeo RR, Leonard JA, Robinson ST, West MR, Mackey AP, Rowe ML, & Gabrieli JD (2018). Beyond the 30-million-word gap: children's conversational exposure is associated with language-related brain function. Psychological Science, 29(5), 700–710. [PubMed: 29442613]

Rosenthal R (1979). The file drawer problem and tolerance for null results. Psychological Bulletin, 86(3), 638–641.

Rosenthal R, & DiMatteo MR (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. Annual Review of Psychology, 52(1), 59–82.

Rowe ML (2012). A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development. Child Development, 83(5), 1762–1774. doi:10.1111/j.1467-8624.2012.01805.x [PubMed: 22716950]

Rufsvold R, Wang Y, Hartman MC, Arora SB, & Smolen ER (2018). The Impact of Language Input on Deaf and Hard of Hearing Preschool Children Who Use Listening and Spoken Language. American Annals of the Deaf, 163(1), 35–60. [PubMed: 29731472]

Sachse S, & Von Suchodoletz W (2008). Early identification of language delay by direct language assessment or parent report? Journal of Developmental & Behavioral Pediatrics, 29(1), 34–41. [PubMed: 18300723]

Salo VC, Rowe ML, & Reeb-Sutherland BC (2018). Exploring infant gesture and joint attention as related constructs and as predictors of later language. Infancy, 23(3), 432–452. [PubMed: 29725273]

Sameroff A (1975). Transactional models in early social relations. Human development, 18(1–2), 65–79.

Schwarzer G, & Schwarzer MG (2012). Package 'meta'. The R Foundation for Statistical Computing, 9.

Soderstrom M, & Wittebolle K (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. PLoS One, 8(11), e80646. [PubMed: 24260443]

Song JY, Demuth K, & Morgan J (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. The Journal of the Acoustical Society of America, 128(1), 389–400. [PubMed: 20649233]

Sperry DE, Sperry LL, & Miller PJ (2019). Reexamining the verbal environments of children from different socioeconomic backgrounds. Child Development, 90(4), 1303–1318. [PubMed: 29707767]

Stoel-Gammon C (2011). Relationships between lexical and phonological development in young children. Journal of Child Language, 38(1), 1–34. [PubMed: 20950495]

Stolcke A, & Droppo J (2017). Comparing human and machine errors in conversational speech transcription. arXiv preprint arXiv:1708.08615.

Suskind DL, Leffel KR, Graf E, Hernandez MW, Gunderson EA, Sapolich SG, … Levine SC (2016). A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study. Journal of Child Language, 43(2), 366–406. [PubMed: 26041013]

Suskind DL, Leffel KR, Hernandez MW, Sapolich SG, Suskind E, Kirkham E, & Meehan P (2013). An exploratory study of "Quantitative Linguistic Feedback" effect of LENA feedback on adult language production. Communication Disorders Quarterly, 34(4), 199–209.

Tamis-LeMonda CS, Bornstein MH, & Baumwell L (2001). Maternal responsiveness and children's achievement of language milestones. Child Development, 72(3), 748–767. [PubMed: 11405580]

Topping K, Dekhinet R, & Zeedyk S (2011). Hindrances for parents in enhancing child language. Educational Psychology Review, 23(3), 413–455.

Trainor LJ, Austin CM, & Desjardins RN (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? Psychological Science, 11(3), 188–195. [PubMed: 11273402]

VanDam M, Ambrose SE, & Moeller MP (2012). Quantity of parental language in the home environments of hard-of-hearing 2-year-olds. Journal of Deaf Studies and Deaf Education, 17(4), 402–420. doi:10.1093/deafed/ens025 [PubMed: 22942314]

VanDam M, & Silbert NH (2016). Fidelity of automatic speech processing for adult and child talker classifications. PLoS One, 11(8), e0160588. [PubMed: 27529813]

Vihman MM, Macken MA, Miller R, Simmons H, & Miller J (1985). From babbling to speech: A re-assessment of the continuity issue. Language, 397–445.

Walle E, & Warlaumont AS (2015). Infant Locomotion, the Language Environment, and Language Development: A Home Observation Study. Paper presented at the CogSci.

Wang Y, Lee CS, & Houston DM (2016). Infant-directed speech reduces English-learning infants' preference for trochaic words. The Journal of the Acoustical Society of America, 140(6), 4101–4110. [PubMed: 28040035]

Weisleder A, & Fernald A (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. Psychological Science, 24(11), 2143–2152. doi:10.1177/0956797613488145 [PubMed: 24022649]

Weizman ZO, & Snow CE (2001). Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. Developmental Psychology, 37(2), 265–279. [PubMed: 11269394]

Williams R, Wang Y, Dilley L, & Houston DM (2019). The effect of auditory noise on processing efficiency in children with normal hearing and children with hearing loss. Paper presented at the 2019 Society for Research in Child Development, Baltimore, Maryland.

Xiong W, Droppo J, Huang X, Seide F, Seltzer ML, Stolcke A, … Zweig G (2017). Toward human parity in conversational speech recognition. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 25(12), 2410–2423.

Xu D, Yapanel U, & Gray S (2009). Reliability of the LENA Language Environment Analysis System in young children's natural home environment. Boulder, CO: LENA Foundation, 1–16.

Xu D, Zhang YW, Mao HM, Xin Y, & Xiao L (2012). The correlation of infants' and toddlers' responsiveness with language and cognitive development in Han-language families. Chinese Journal of Evidence-Based Pediatrics, 7(6), 435–439.

Yoder P, & Symons F (2010). Observational measurement of behavior. New York: Springer Publishing Company.

Zimmerman FJ, Gilkerson J, Richards JA, Christakis DA, Xu D, Gray S, & Yapanel U (2009). Teaching by listening: The importance of adult-child conversations to language development. Pediatrics, 124(1), 342–349. [PubMed: 19564318]

Highlights

- This research assessed the predictability of LENA's automated measures for child language
- We showed a medium association between LENA's automated measures and language
- Conversational turn and child vocalization showed medium associations with other measures of child language
- Adult word count showed a small-to-medium association with language

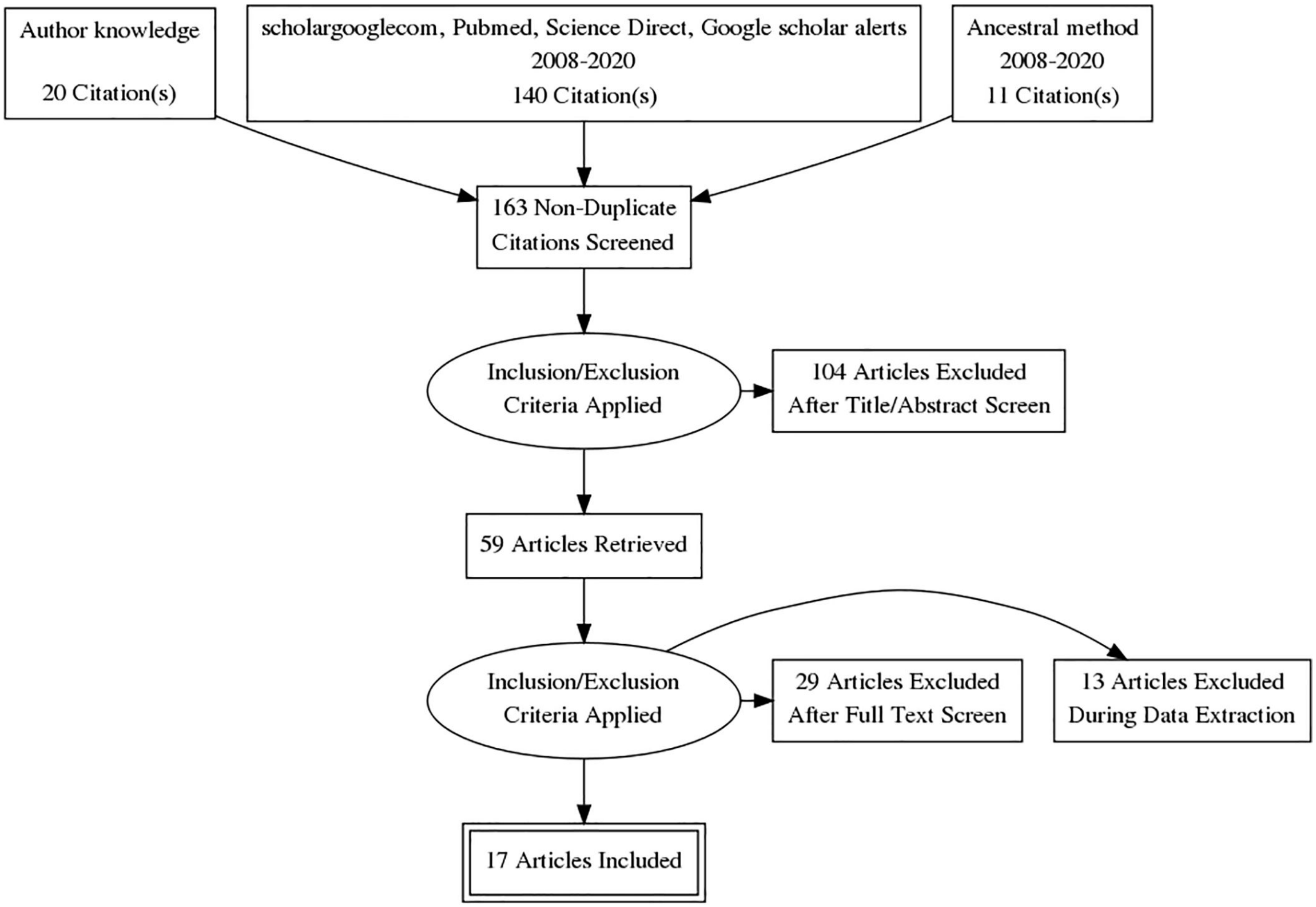

Figure 1.
Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) flow diagram

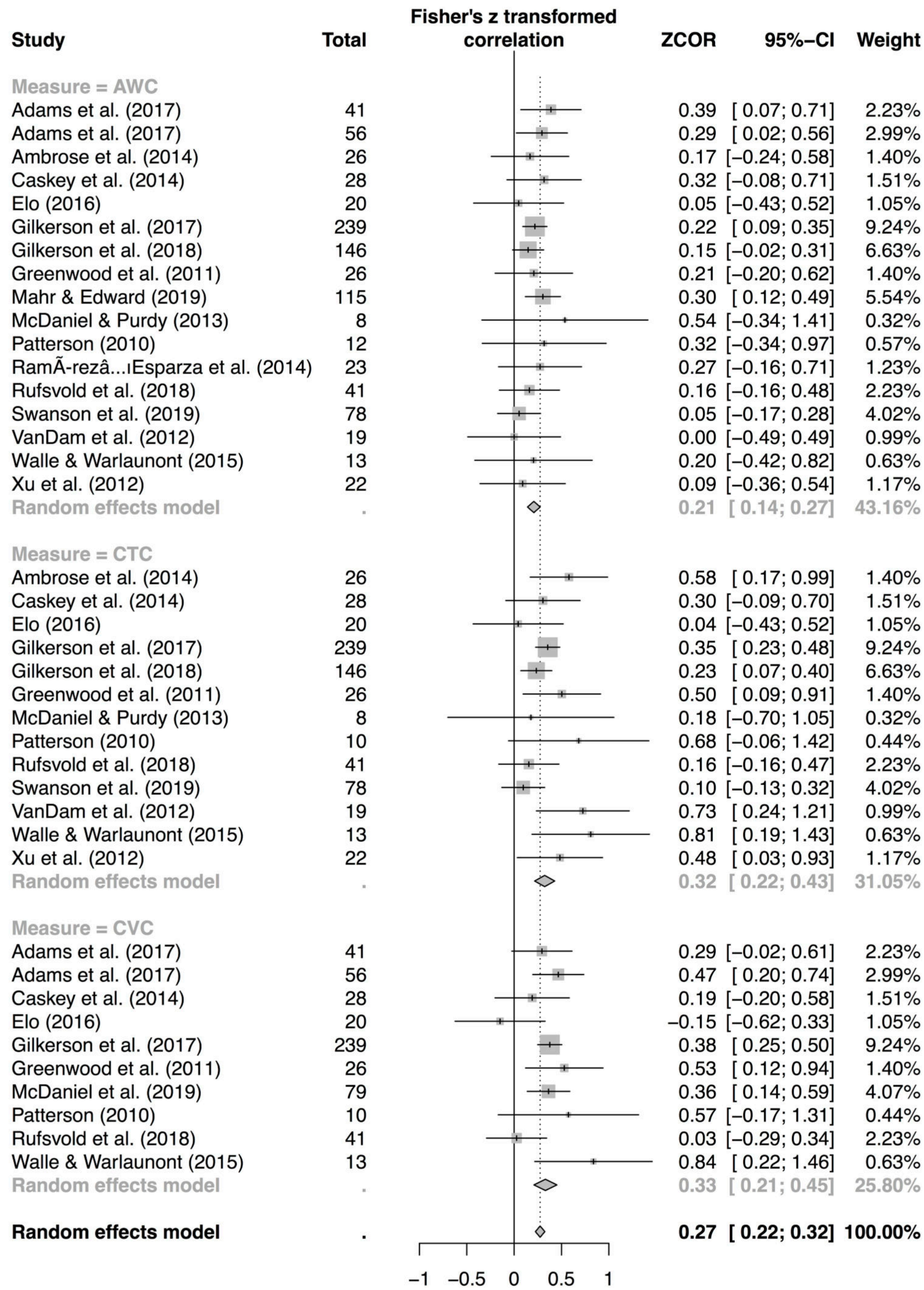

Figure 2.

Forrest plot showing the results of the 17 reports with 40 effect sizes examinig the assocations between LENA's automated measures, AWC, CTC, and CVC and language skills. Total: weighted number of children contributed to the correlation.

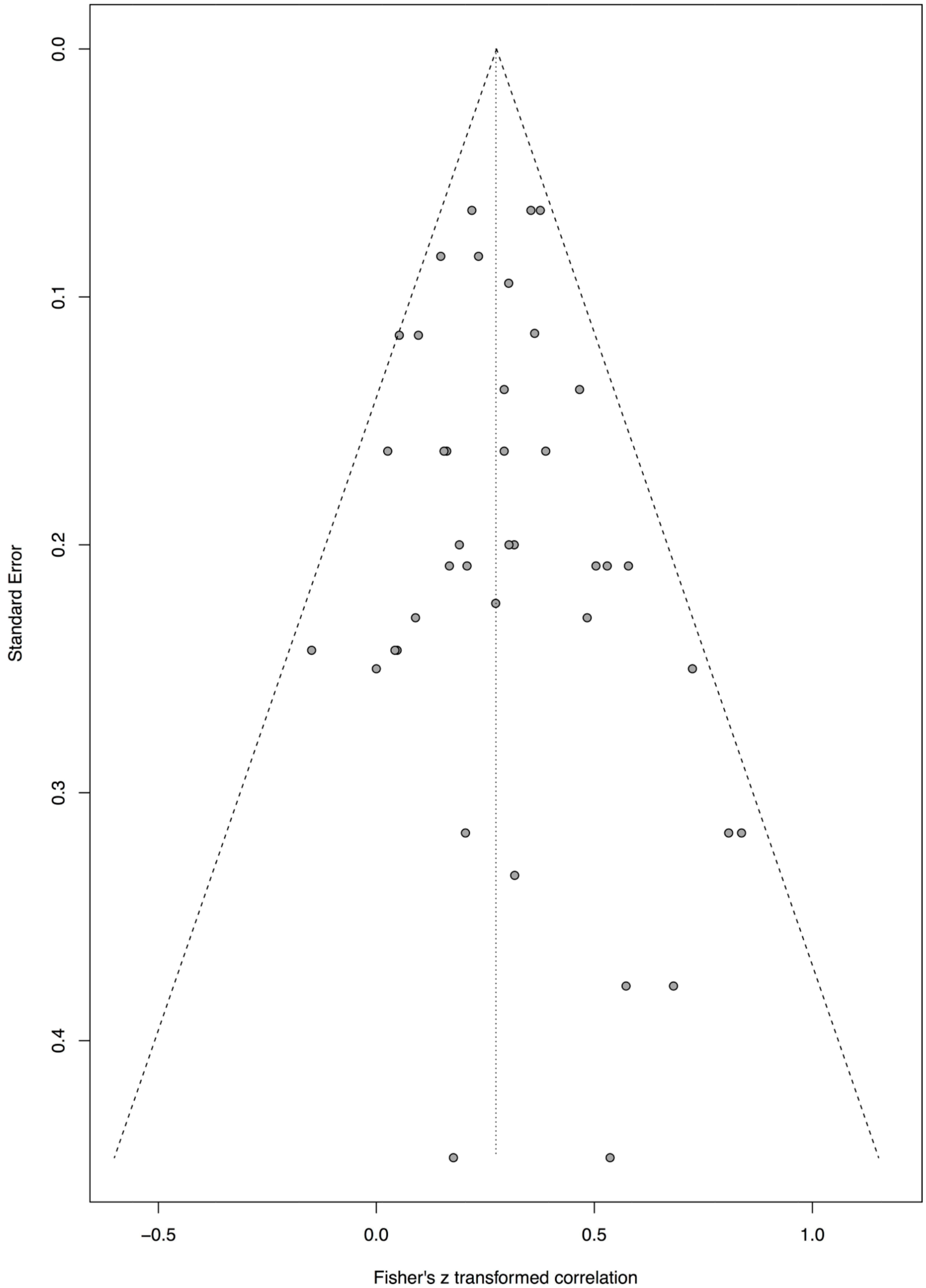

Figure 3.
Funnel plot showing potential publication bias of the included 17 reports with 40 effect sizes reported. The vertical line indicates the overall pooled effect size. On the ordinate, the standard error of each study is shown and on the abscissa, the effect size of each study analyzed is shown in Fisher's z units. The circles plot each study on the funnel plot with higher publication bias indicated by circles outside the "funnel."

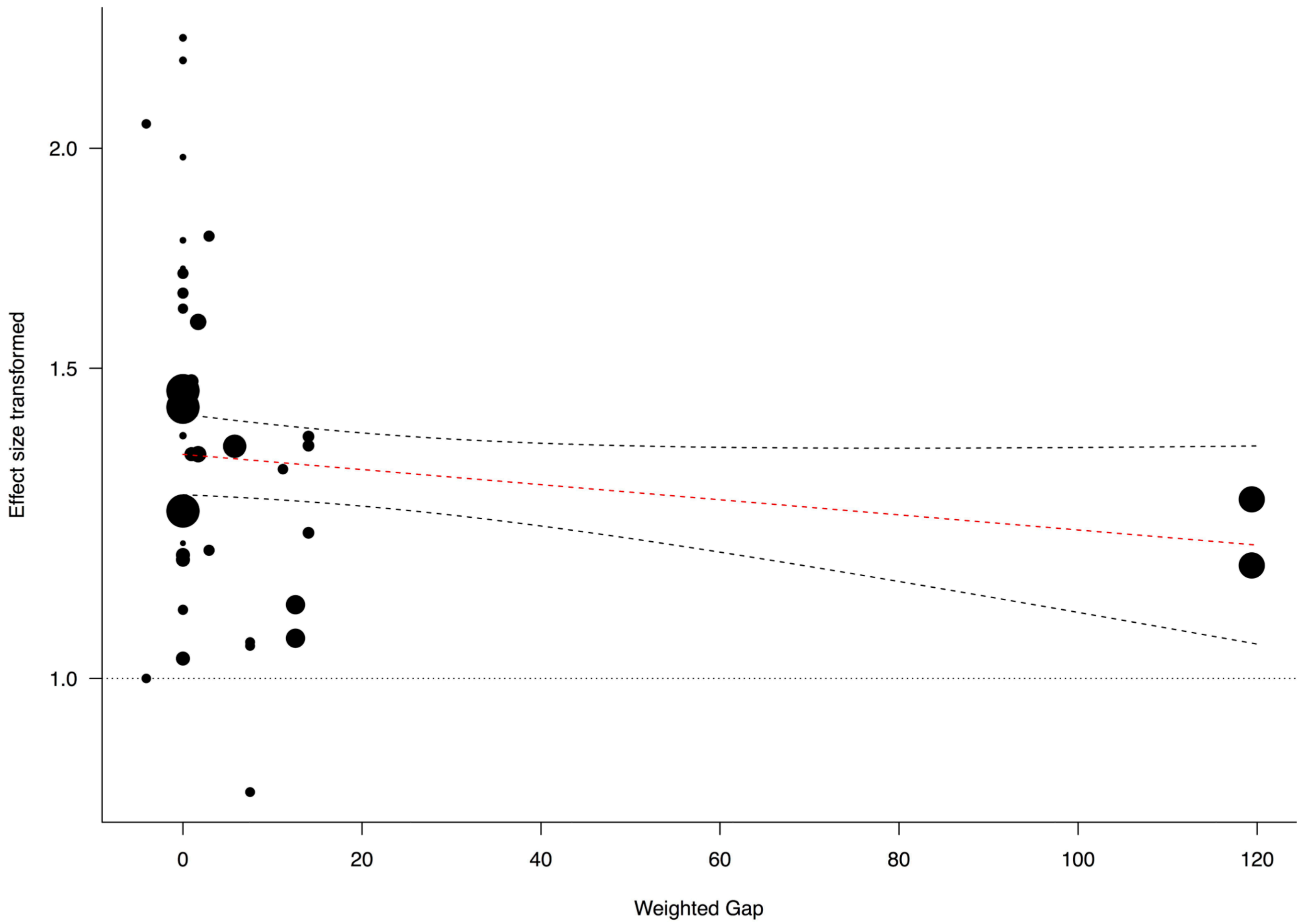

Figure 4.

Scatterplot showing the observed correlation (exponentially transformed) of individual effect size plotted against the quantitative predictor, the weighted gap between LENA recording and language test based on mixed-effects model. The radius of the points is drawn proportional to the inverse of the standard errors: larger/more precise studies are shown as larger points.

Author Manuscript
Author Manuscript
Author Manuscript

Table 1.

Characteristics of the included report

Study index		Pub status	Child No.	Country	Primary language	Dev Status	Mean age at LENA (month)	LENA's automated measures	No. of recordings	Duration of LENA recording (Hr)	Age at language measures (month)	Outcome measure	Language test modality	Language test method
1	Adams et al. (2018)	1	56	US	English	TD	16.3	AWC CVC	1	14.7	18	Bayley, MCDI	Mixed	Mixed
1	Adams et al. (2018)	1	41	US	English	Preterm	16.1	AWC CVC	1	14.7	16.1 18	Bayley, MCDI	Mixed	Mixed
2	Ambrose et al. (2014)	1	28	US	English	HL	25.75	AWC CTC	6.14	12.19	24.7 36.6	MSEL, CASL	Mixed	Mixed
3	Caskey et al. (2014)	1	36	US	English	Preterm	−1.59 −0.85	AWC CTC CVC	2	16	7 18	Bayley	Mixed	Direct child assessment
4	Elo (2016)	2	20	Finland	Finnish	Mixed[a]	9	AWC CTC CVC	7.1	10.56	12 18 24	MCDI	Mixed	Parent report
5	Gilkerson, Richards, Warren, et al. (2017)	1	306	US	English	TD	25	AWC CTC CVC	9.8	12	Concurrent	PLS, REEL, MCDI, CDI	Expressive	Mixed
6	Gilkerson et al. (2018)	1	146	US	English	TD	21.48	AWC CTC	5.17	12	140.88	VCI, EVT, PPVT	Mixed	Direct child assessment
7	Greenwood et al. (2011)	1	30	US	English	TD	15.6	AWC CTC CVC	24	12	Concurrent	PLS	Mixed	Direct child assessment
8	R. McDaniel and Purdy (2013)	2	8	New Zealand	English	HL	40	AWC CVC	3	8–12	Concurrent	PLS	Mixed	Direct child assessment
9	Mahr & Edward (2019)	1	121	US	English	TD	32.9	AWC	2	>10	38.68	EVT, PPVT	Mixed	Direct child assessment
10	J. McDaniel, Yoder, Estes, and Rogers (2019)	2	87	US	English	ASD	23.42 29.42	CVC	NR	NR	Concurrent Longitudinal (6 months interval)	MCDI, MSEL, VABS, CSP	Expressive	Direct child assessment
11	Patterson (2010)	2	12	Canada	English	ASD	33	AWC CTC CVC	NR	9.5	Concurrent	MCDI	Mixed	Parent report
12	Ramírez-Esparza, García-Sierra,	1	26	US	English	TD	12.84	AWC	4	8	24	MCDI	Expressive	Parent report

	Study index	Pub status	Child No.	Country	Primary language	Dev Status	Mean age at LENA (month)	LENA's automated measures	No. of recordings	Duration of LENA recording (Hr)	Age at language measures (month)	Outcome measure	Language test modality	Language test method
	and Kuhl (2017)													
13	Rufsvold, Wang, Hartman, Arora, and Smolen (2018)	1	41	US	English	Mixed[b]	46.93	AWC CTC CVC	1	16	Concurrent	PPVT	Receptive	Direct child assessment
14	Swanson	1	78	US	English	Mixed[c]	12.77	AWC	2	15.77	25.34	Mullen, MSEL	Mixed	Direct child assessment
15	VanDam et al. (2012)	1	22	US	English	HL	29.4	AWC CTC	1	11.89	25.3	MSEL	Receptive	Mixed
16	Walle and Warlaumont (2015)	2	13	US	English	Mixed[d]	12.75	AWC CTC CVC	1	15.23	Concurrent	MCDI	Mixed	Parent report
17	Xu et al. (2012)	1	22	China	Mandarin	TD	16.80	AWC CTC	3	16	Concurrent	ILDSS	Mixed	Mixed

Note: Pub Status: Publication status: 1 (peer-reviewed published journal articles), 2 (non-peer-reviewed or unpublished reports); Child No.: the number of children who participated in the studies; Dev Status: developmental status. TD: children with typically-developing; HL: children with hearing loss; ASD: children with Autism Spectrum Disorders; Mixed[a]: the population included preterm and term infants; Mixed[b]: the population included both children with NH and children with HL; Mixed[c]: the population include children with ASD and TD; Mixed[d]: the population included walking and crawling infants; AWC: adult word count, CTC: conversational turn count; CVC: child vocalization count; NR indicated that “data missing or not reported”. Bayley: the Bayley Scales of Infant and Toddler Development; CDI: Child Development Inventory; CASL: the Comprehensive Assessment of Spoken Language; CSP: Communication Sample Procedure; EVT: Expressive Vocabulary Test; ILDSS: Infant Language Development Screening Scales; MCDI: the MacArthur-Bates Communicative Development Inventory; MSEL: Mullen Scale of Early Learning: PLS: the Preschool Language Scale; PPVT: Peabody Picture Vocabulary Test; REEL: the Receptive Expressive Emergent Language Test; VABS: Vineland Adaptive Behavior Scale; VCI: Verbal Comprehension Index.

Table 2.

Meta-analytic results of moderator table

Moderator variables	No. of participants	No. of effect sizes	Effect size (r)	95% CI	Q	I^2	*p*
Categorical							
Child developmental status					1.17		.759
ASD	99	4	0.38***	[.20, .53]	.93	.00	
HL	58	6	0.35**	[.12, .55]	6.47	.23	
Preterm	77	5	0.30***	[.15, .43]	.61	.00	
TD	707	14	0.29***	[.24, .34]	12.86	.00	
Publication status					.28		.595
Peer-reviewed	954	30	.26***	[.22, .31]	30.96	.06	
Non-peer-reviewed	139	10	.31**	[.12, .47]	13.00	.31	
Language test modality					2.55		.110
Receptive	60	5	.18^{+}	[−.03, .37]	6.19	.35	
Expressive	413	5	.31***	[.25, .37]	3.62	.00	
Language test method					.43		.511
Direct child assessment	545	17	.21***	[.15, .27]	13.07	.00	
Parent report	78	10	.29**	[.09, .47]	12.70	.291	
Continuous							
Age at LENA recording					.00		.984
Gap between LENA recording and language test					2.66		.103^{+}

*** $p < .001$

** $.001 < p < .01$

* $.001 < p < .05$

$^{+}$ $.05 < p < .10$