Research Article

Expressive Language in Preschoolers Born Preterm: Results of Language Sample Analysis and Standardized Assessment

Caitlin M. Imgrund,^a Diane F. Loeb,^b and Steven M. Barlow^c

Purpose: Preschoolers born preterm are at an increased risk for the development of language impairments. The primary objective of this study was to document the expressive language skills of preschoolers born preterm through 2 assessment procedures, language sample analysis, and standardized assessment. A secondary objective was to investigate the role of nonlinguistic factors in standardized assessment scores.

Method: The language skills of 29 children born preterm (mean gestational age of 31 weeks) were compared to a group of 29 preschoolers born full term. Language samples were collected during free play and objective measures of semantic and grammatical skills were calculated. Likewise, grammatical and semantic measures of language were obtained from a standardized assessment. Information on nonlinguistic factors, including attention, hyperactivity, and nonverbal intelligence quotient, was also collected.

Results: The results of analyses of variance indicated that the children in the PT group had significantly poorer performance than the children born full term on all of the measures of language skill obtained from the language sample analysis. Group differences were found for only 1 measure of language skill obtained from the standardized assessments. Nonverbal factors were not found to account for group differences in assessment scores.

Conclusions: Generally, the children born preterm performed more poorly when language skill was measured via language sample analysis than standardized assessment. These findings underscore the importance of using language sample analysis in conjunction with standardized assessment in the evaluation of children's developing language skills.

Preterm birth, or birth that occurs before 37 completed weeks of gestation, is a significant medical issue in the United States. Around one in 10 children are born preterm, which results in approximately half a million preterm births annually (Martin, Hamilton, Osterman, Driscoll, & Matthews, 2017). Advances in the care of these children in neonatal intensive care units (NICUs) have resulted in improved survival rates since the late 1980s (Paranka, Yoder, & Brehm, 1999; Polin, Carlo, & Committee on Fetus and Newborn, 2014). Children who

were born preterm are at an increased risk for adverse outcomes across a broad spectrum of neurodevelopmental domains (Bos, Van Braeckel, Hitzert, Tanis, & Roze, 2013; Mangrin, Horwood, & Woodward, 2017; Zimmerman, 2017). Indeed, these children constitute one of the largest groups of children at risk for poor neurodevelopmental outcomes.

Recent research has been conducted to better understand the impact of preterm birth on neurodevelopmental outcomes and trajectories. A meta-analysis of 27 studies conducted to determine the association between preterm birth and subsequent cognitive abilities found that children born preterm had intelligence quotient (IQ) scores that were, on average, 11.94 points lower than their peers who were born full term (Kerr-Wilson, Mackay, Smith, & Pell, 2012). These cognitive deficits may impact school performance and subsequent learning abilities, with as many as 65% of children born at or before 32 weeks of gestational age (GA) identified with learning disabilities (Grunau, Whitfield, & Davis, 2002). The attentional skills of children born preterm have also been the subject of research. A meta-analysis of

^aDepartment of Communication Sciences and Disorders, Florida Atlantic University, Boca Raton

^bDepartment of Communication Sciences and Disorders, Baylor University, Waco, TX

^cDepartment of Special Education and Communication Disorders, University of Nebraska–Lincoln

Correspondence to Caitlin M. Imgrund: cimgrund@fau.edu

Editor-in-Chief: Sean Redmond

Editor: GERALYN TIMMER

Received June 4, 2018

Revision received October 1, 2018

Accepted October 12, 2018

https://doi.org/10.1044/2018_JSLHR-L-18-0224

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

these studies indicated that children who were born preterm were at an increased risk for attention-deficit/hyperactivity disorder as well as other behavioral problems (Bhutta, Cleves, Casey, Cradock, & Anand, 2002).

As with other areas of neurodevelopment, the language skills of children born preterm are also adversely impacted. Studies of language development in this population have documented language deficits from infancy through adolescence, and even into adulthood (Casiro et al., 1990; Foster-Cohen, Edgin, Champion, & Woodward, 2007; Magill-Evans, Harrison, Van der Zalm, & Holdgrafer, 2002; van Noort-van der Spek, Franken, & Weisglas-Kuperus, 2012; Zimmerman, 2017).

The preschool period is an important time for language development; consequently, several studies have investigated the rates of language impairment of preschoolers who were born preterm. Foster-Cohen, Friesen, Champion, and Woodward (2010) found high prevalence of mild-to-moderate language impairment in 4-year-old children who were born preterm. In their sample of 100 children born preterm, 16% had a mild language delay (1–1.5 *SDs* below the mean on the Clinical Evaluation of Language Fundamentals Preschool–Second Edition [CELF Preschool-2]; Wiig, Secord, & Semel, 2004) and 15% had a moderate language delay (1.5–2 *SDs* below the mean on the CELF Preschool-2) in comparison to 8.6% and 6.7%, respectively, in the full-term (FT) comparison group. Sansavini et al. (2010) also investigated the rates of language impairment in children born preterm but without significant cerebral damage at age 3;6 (years;months). Their sample had similar rates of language impairment, with approximately one third (34.4%) of the children in the preterm (PT) group demonstrating significant delays in language acquisition in comparison to 7.5% of children in the FT comparison group. Preschoolers who were born preterm and are displaying deficits in language acquisition are not likely to catch up to their full-term peers, as language scores in this population tend to be fairly stable over time (Putnick, Bornstein, Eryigit-Madzwamuse, & Wolke, 2017).

Because children born preterm are at an increased risk for adverse language outcomes, the accurate assessment of language skills in this population is particularly crucial in order to identify children who are in need of early intervention services. Researchers and clinicians have a number of methods available to aid them in this task. Two of the most commonly used methods for the assessment of a child's language skills are standardized assessment and language sample analysis. Standardized assessments of language ability offer clinicians and researchers alike an efficient means for assessing a broad array of linguistic skills in a relatively short period of time. Because many standardized assessments are also norm referenced, these types of assessments are highly useful for determining if a child's language abilities differ significantly from his or her same-age peers. Indeed, the use of norm-referenced assessments is a core component of what is considered the “gold standard” in the assessment of language (Tomblin, Records, & Zhang, 1996). The use of standardized assessment results, especially as a stand-alone indicator of language ability, is not without its drawbacks.

Some researchers have called into question the ecological validity of standardized assessments (Costanza-Smith, 2010; Marquardt & Gillam, 1999). Additionally, these types of assessments may not be sensitive enough to detect language impairments in some populations of children.

Language sample analysis is also a highly relevant assessment tool in both clinical and research contexts. A language sample offers a rich and robust source of data from which a wide array of analyses can be conducted, including analyses of the expressive use of semantics, morphology, syntax, and pragmatics. The data obtained from the analysis of language samples may be more accurate for the diagnosis of language impairment in children than standardized assessments. Aram, Morris, and Hall (1993) investigated the overlap between assessment measures and clinical diagnoses of language impairment. The authors found that mean length of utterance (MLU) obtained from a language sample was a more sensitive measure than the standardized assessment results for positively identifying children who had a clinical diagnosis of language impairment. A follow-up analysis conducted by Dunn, Flax, Sliwinski, and Aram (1996) found that children who had been clinically diagnosed with language impairment but were not identified as impaired through standardized assessment had significant difficulty with language measures obtained through language sample analysis.

Although studies of language development in a particular population of children would ideally include information obtained from both standardized assessment and language sample analysis, studies of the language outcomes of children who were born preterm have largely employed the use of standardized, norm-referenced assessments. Despite the diagnostic utility of language sample analysis as a means for assessing the linguistic systems of children, very few studies on the language outcomes of children born preterm have analyzed language samples in conjunction with standardized assessment outcomes. Mahurin Smith, DeThorne, Logan, Channell, and Petrill (2014) analyzed the results of both language sample analysis and standardized assessments in 57 children born preterm and the comparison group of full-term peers when the children were between 7 and 10 years of age. In their sample, the children who were born preterm performed more poorly than the children in the FT comparison group on the majority of language variables obtained from standardized assessment. Interestingly, no group differences were seen for the language variables obtained from the language samples. In a follow-up study, Mahurin-Smith, DeThorne, and Petrill (2017) found that attention skills, as measured by a parent report of attention skills, partially mediated the relationship between prematurity and outcomes on the standardized assessment, suggesting that poor attention skills may have accounted for poorer performance on the standardized assessments.

Crosbie, Holm, Wandschneider, and Hemsley (2010) found a comparable pattern. In a sample of school-age children born preterm, the children born preterm exhibited poorer performance than the children in the FT comparison group on standardized assessments of language, but no

group differences were seen for most measures of language skill obtained from a narrative language sample (e.g., MLU, number of different words). These two studies provide emerging evidence that children who were born preterm may perform more poorly than their full-term counterparts when language is measured through standardized assessment rather than via language sample analysis. These results are counterintuitive to conventional wisdom and previous studies that language sample analysis is more sensitive to identifying language difficulties in children. However, it may be that the children in the latter studies had a comorbidity of language and attentional problems, which may have negatively impacted their standardized test results, but not language sampling results. It may also be that these results were mediated by the children's ages and that the older the child becomes, the less sensitive language sampling measures are.

The primary purpose of this study was to investigate the impact of preterm birth on language outcomes in preschoolers as measured by both standardized assessment and language sample analysis. Language sample analysis provides valuable information and has been underutilized as a tool for assessing the language skills of this population. A secondary aim of this study was to explore the impact of nonlinguistic factors, such as attention, hyperactivity, and nonverbal intelligence, on the performance of standardized language assessments.

The following research questions were addressed in this study.

1. Do the semantic and syntactic skills of preschool-age children born preterm differ from those of full term, typically developing children as measured by standardized assessment and language sample analysis?
2. Do the nonlinguistic factors of attention, hyperactivity, and nonverbal IQ of preschool-age children born preterm differ from those of full term, typically developing children?
3. Are group differences on standardized assessments of language exhibited after controlling for the effects of the nonlinguistic factors of attention, hyperactivity, and nonverbal intelligence?

On the basis of previous research (Crosbie, Holm, Wandschneider, & Hemsley, 2010; Mahurin Smith et al., 2014), we predicted that the children in the PT group would perform more poorly than the children in the FT group when language was measured through standardized assessment, but not language sample analysis. We also predicted that group differences would be evident for the nonlinguistic factors of interest and that these factors would negate any group differences exhibited on the standardized assessments.

Method

Participants

The children in the PT group were recruited from an existing database of 92 children who were participants in a study of neurodevelopmental outcomes following

patterned orocutaneous stimulation while the children were hospitalized in the NICU (Loeb, Imgrund, Lee, & Barlow, 2018). Inclusionary criteria for this study were birth GA before 36 weeks as determined by obstetric ultrasound and clinical examination, tube feedings while in the NICU, head circumference within the 10th–90th percentile of mean for post menstrual age (PMA), neurological examination showing no anomalies, response to light and sound, spontaneous movements of all extremities, and stable vital signs. Infants with intracranial hemorrhage Grades III and IV, neonatal seizures, periventricular leukomalacia, necrotizing enterocolitis, meningitis, sepsis, chromosomal anomalies, or craniofacial malformation were excluded from the study. Thirty children born preterm were recruited to participate in the study. One child recruited for the PT group was unable to complete the assessments due to a significant disability caused by both her preterm birth and exposure to neurotoxic chemotherapy treatments early in life. Removal of this child from the data set resulted in a total of 29 children in the PT group (18 boys, 11 girls).

The children were between the ages of 4;0 and 4;11 at the time of their participation. GA at birth ranged from 23 to 34 completed weeks of gestation, with a mean GA of 31 weeks ($SD = 18$ days). The birth weights of the children in this group ranged from 610 to 2,340 g, with a mean birth weight of 1,507 g ($SD = 461$ g). Socioeconomic status was measured in this study through maternal education level. The mothers of the children who participated in this study came from a variety of education levels. English was the primary language spoken in the home for all participants.

A typically developing FT comparison group was recruited from the community via flyers distributed to local preschools and community posting boards. The PT and FT groups contained similar gender and racial/ethnicity ratios and had comparable levels of maternal education, which was used as a proxy measure of socioeconomic status. The two groups did not differ with respect to age at assessment. Only children with nonverbal IQ scores and composite language scores above 85 were considered eligible for the FT comparison group. Additional exclusionary criteria included birth before 37 completed weeks of gestation; diagnosed neurodevelopmental, sensory, or behavioral disorder; and history of language, speech, or hearing impairment (see Table 1).

The participants attended a single assessment session, which took place at a quiet location within the community, such as a room in a local library or clinic. The primary investigator ensured that the room was free from distractions. Parents were permitted to attend the session with their child but were asked to refrain from interacting with the child while the assessments were being administered. The sessions were approximately 90 min in length, including time for short breaks in between assessments and a longer snack break. The administration order of the assessments was randomized to prevent order effects from influencing the assessment results. The Human Subjects Committee of the Institutional Review Board at the University of Kansas approved the research protocol;

Table 1. Participant characteristics.

Variable	Preterm		Full term	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age (years;months)	4;5	3.1 (months)	4;4	3.5 (months)
Gestational age (weeks)	31	18 (days)		
Birth weight (g)	1,507.86	461		
	Preterm		Full term	
	<i>n</i>	%	<i>n</i>	%
Gender				
Male	18	62.1	16	55
Female	11	37.9	13	44
Maternal education				
Some high school or less	1	3.4	0	0
High school/GED	0	0	1	3.4
Some college/associate's degree	9	31	9	31
Bachelor's degree	11	37.9	12	41.4
Graduate degree	8	27.6	7	24.1
Race				
White	26	89.6	24	82.8
Black/African American	1	3.4	2	6.9
Asian	0	0	0	0
American Indian/Alaskan Native	0	0	0	0
More than one race	2	6.9	3	10.3
Ethnicity				
Hispanic/Latino	3	10.3	5	17.2
Not Hispanic/Latino	26	89.6	24	82.8

Note. GED = General Educational Development.

informed consent was obtained from the parents, and assent was obtained from the children prior to participation.

Assessment of Language

The CELF Preschool-2 (Wiig, Secord, & Semel, 2004), a psychometrically strong assessment tool with high levels of both reliability and validity, was administered to all participants. Although the assessment was administered in its entirety for eligibility purposes, only the expressive semantic and grammatical subtests were pertinent to the aims of this study. These subtests were (a) Word Structure, (b) Expressive Vocabulary, and (c) Recalling Sentences. The Word Structure subtest was designed to evaluate a child's use of grammatical rules in a sentence completion task. The Expressive Vocabulary subtest was designed to measure expressive semantics and requires the child to label pictures of people, objects, and actions. Finally, the Recalling Sentences subtest was designed to evaluate a child's ability to repeat sentences of increasing length and grammatical complexity verbatim.

A language sample was obtained from each of the participants in order to assess the conversational language skills of the children through language sample analysis. Language samples were obtained during a period of free play with the examiner. A standard set of toys was provided. The toys were presented to the child in a randomized order, and three sets of toys were utilized (a toy barn set with accompanying animals, two baby dolls with a diaper bag, and a set of fairytale-themed puppets including a dragon, prince,

princess, and knight). These items were selected because of their high interest level for preschool-age children and their ability to provide appropriate contexts for the elicitation of a variety of linguistic structures of interest. During the collection of the language sample, Leadholm and Miller's (1992) recommendations of following the child's lead and using primarily open-ended prompts and questions were employed. The language samples were video- and audio-recorded.

The language samples were transcribed by the primary investigator using the standard conventions of the Systematic Analysis of Language Transcripts (Miller & Chapman, 2007). So that each analysis was of an equal length, only 100 complete and intelligible utterances were used in the analyses, which is a commonly used standard in the analysis of language samples (Cole, Mills, & Dale, 1989; Heilmann, 2010; Miller, 1981).

Measures of semantic and grammatical ability were derived from the language samples. Semantic ability was measured through two variables: (a) number of different word roots (NDW) and (b) a semantic analysis score. NDW has been used as an assessment measure to distinguish between children with and without typical language abilities and is considered to be a valid measure of lexical diversity (Malvern & Richards, 2012; Watkins, Kelly, Harbers, & Hollis, 1995). A correlational study found NDW to be more strongly correlated with scores obtained on standardized assessments of vocabulary than other measures of semantic ability derived from language samples such as type-token ratio (Ukrainetz & Bloomquist, 2002). Additionally,

a semantic analysis score was computed for each of the samples. This procedure was adapted from Mahurin Smith et al. (2014) and was included to quantify the child's use of less frequent, non-early-acquired words. This score is thought to reflect a child's ability to utilize more nuanced, mature vocabulary, for example, using the word "disgusting" instead of "yucky." In order to conduct this analysis, each of the different word roots spoken by the individual child was recorded in a list and tallied. From this list of word roots, a series of high-frequency words commonly used by young children were removed. High-frequency words were identified by using the MacArthur-Bates database of words produced commonly by toddlers (Fenson et al., 1993). Proper nouns, sound effects, and numbers were likewise removed from the list. The remaining lower frequency words were then tallied for each child, resulting in a semantic analysis score.

In order to assess conversational grammatical abilities, three distinct measures of syntactic and morphological skills were obtained from the language samples: (a) MLU, (b) percentage correct use of finite verb morphology in obligatory contexts, and (c) developmental sentence score (DSS). These measures were selected for their ability to assess different aspects of grammatical development.

The first measure, MLU in morphemes, was calculated for each of the language samples using the Systematic Analysis of Language Transcripts software. MLU is commonly used as a gross measure of syntactic ability and, when used in conjunction with other measures, is considered a valuable gauge of a child's productive language skills (Eisenberg, Fersko, & Lundgren, 2001). The second grammatical measure was based on the children's production of finite verb morphology in obligatory contexts. Because difficulty with finite verb morphology is a key indicator of language impairment in young children, the assessment of finite verb morphology is considered a particularly robust assessment measure (Goffman & Leonard, 2000; Ingram, 1974; Rice & Wexler, 2001). Goffman and Leonard's (2000) methods for calculating percentage correct use of finite verb morphology in obligatory contexts were utilized in this study. The finite verbs and morphemes under investigation included past tense *-ed*, third-person singular *-s*, and the copula and auxiliary verbs *is*, *are*, *am*, *was*, and *were*. *Do* and *have* were not included in the calculations per the methods of Goffman and Leonard (2000), who noted that these forms are not frequently used by children in this age range and because of potential difficulty in reliably noting obligatory contexts of these verbs in spontaneous language samples. The percentage of correct usage in obligatory contexts was calculated by dividing the total number of correct productions by the total number of obligatory contexts and multiplying by 100. During the language sample, the examiner periodically used prompts to elicit an adequate number of obligatory contexts for third-person present *-s* and past tense *-ed*. For example, the researcher used the prompt "tell me what your ____ (e.g., *baby*, *car*, *boy*) does?" (adapted from Rice & Wexler, 2001) to elicit third-person singular *-s*. To attempt to elicit regular past tense *-ed*, the examiner periodically focused the play on actions that are

conventionally named in English with a regular verb (e.g., *walk*, *hug*, *kiss*, *play*, *jump*) and prompted with "what happened?" (adapted from Rice & Wexler, 1996).

The third grammatical measure obtained from the language sample was the child's DSS (Lee, 1974). DSS is a procedure for quantifying a child's syntactic (and, to a minimal extent, semantic) abilities. The DSS is computed on the basis of eight syntactic categories: indefinite pronouns, personal pronouns, main verbs, secondary verbs, negatives, conjunctions, interrogative reversals, and *wh*-questions. Each utterance is given a point value ranging between 1 and 8 for each of the syntactic categories used in that utterance. Early-acquired syntactic constructions are assigned lower point values, and later-acquired syntactic constructions are assigned higher point values. Only utterances that are complete, intelligible, and unique are included in the final analysis, and at least 50% of the total utterances contained in the sample must be complete, intelligible, and unique in order for a DSS to be calculated. DSS provides extensive information about a child's developing syntactic system, and despite being over 40 years old, DSS maintains its utility as a measure of syntactic ability in both clinical and research contexts (Hughes, Fey, & Long, 1992). DSS has also been shown to effectively differentiate between children with and without language impairments (Paul, 1996).

Transcription and coding reliability analyses were conducted by a second, independent researcher in order to ensure that the language samples were reliably transcribed and that the language sample variables were reliably calculated. Twenty percent of the PT and FT samples were transcribed independently by a second researcher without access to the original transcript. Transcription reliability was calculated by comparing the two transcripts on a morpheme-by-morpheme basis and determining the number of agreements and disagreements. Interrater reliability for the transcription of the language samples ranged from 92.3% to 97.4% ($M = 94.7\%$) in the PT group and from 94% to 96.3% ($M = 95.4\%$) in the FT group. The point-to-point reliability scores for the variables of interest ranged from 87.5% to 99.3%, indicating that the data obtained from the language samples were of adequate reliability.

Nonlinguistic Measures

In addition to measures of language, several variables addressing the nonlinguistic developmental skills of nonverbal intelligence, attention, and hyperactivity were assessed. Nonverbal intelligence was assessed using the Matrices subtest of the Kaufman Brief Intelligence Test–Second Edition (Kaufman & Kaufman, 2004). Two measures of attention were utilized: the Attention subscale of the Behavior Assessment System for Children–Second Edition (BASC-2; Reynolds & Kamphaus, 2015) and the Conners Kiddie Continuous Performance Test–Second Edition (Conners K-CPT 2; Conners, 2015). The BASC-2 is a parent report measure, and the Conners K-CPT 2 is a computerized, performance-based measure of attention. For this assessment, the child was presented a series of pictures and asked to press a button each time a picture

appeared but refrain from pressing the button if the picture was a soccer ball. For the purposes of this study, only the omission *T* scores were utilized, which are a measure of inattentiveness (Conners, 2015). Omission *T* scores directly reflect the number of targets not responded to by the child. Hyperactivity was measured through scores obtained on the BASC-2 Hyperactivity subscale.

Results

Descriptive statistics for the variables of interest can be found in Table 2. To determine if differences between the PT and FT groups were present for the variables of interest, a series of analyses of variance (ANOVAs) were conducted to compare the performance of the PT group to that of the FT group on each of the variables obtained from the language samples, the expressive subtests of the CELF Preschool-2, and the nonlinguistic factors. A modified Bonferroni adjustment was used to determine the statistical significance of the *p* values for all sets of ANOVAs. This adjustment was used in order to preserve power while ensuring the probability of committing a Type I error remained below .05.

Standardized Assessment of Language Results

The children in the PT group achieved a mean Core Language standard score of 101.83 on the CELF Preschool-2 (*SD* = 10.43). The mean preterm Expressive Language Index was 100.17 (*SD* = 11.08), and the mean Receptive Language Index was 98.00 (*SD* = 11.46). The children in the FT group achieved a mean Core Language standard score of 104.34 (*SD* = 7.98). The mean full-term

Expressive Language Index was 104.07 (*SD* = 8.18), and the mean Receptive Language Index was 103.93 (*SD* = 9.29). A one-way between-subjects ANOVA was conducted to compare performance on the expressive language subtests of the CELF Preschool-2, which include one subtest assessing expressive semantics (i.e., Expressive Vocabulary subtest) and two subtests assessing expressive grammar (i.e., Word Structure and Recalling Sentences subtests). For the semantic measure, there was not a statistically significant difference between the PT and FT groups, $F(1, 56) = 0.256$, $p = .617$. The magnitude of the effect was negligible ($d = 0.135$). For the grammatical measures, there was not a significant difference between groups on performance on the CELF Preschool-2 Word Structure subtest, $F(1, 56) = 1.88$, $p = .176$, with a small effect size ($d = 0.366$). Statistically significant group differences were seen on the CELF Preschool-2 Recalling Sentences subtest, $F(1, 56) = 7.99$, $p = .007$. The magnitude of the effect was medium ($d = 0.755$), indicating both statistical and practical significance for this measure (see Table 3).

Language Sample Analysis Results

Likewise, a one-way between-subjects ANOVA was conducted to compare performance between the FT and PT groups on the measures obtained from language sample analysis. Statistically significant group differences were found for all measures obtained from the language sample analyses. For the semantic measures, significant group differences were found for both NDW, $F(1, 56) = 7.612$, $p = .008$, and the semantic analysis score, $F(1, 56) = 6.883$, $p = .011$. The magnitudes of the effects were medium ($d = 0.754$ and $d = 0.70$, respectively). On the grammatical measures,

Table 2. Descriptive statistics for dependent variables of interest.

Variable	Preterm		Full term	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Language sample analysis				
Semantic: NDW	130.34	19.34	142.86	14.93
Semantic: Semantic analysis score	35.90	9.54	41.72	7.21
Grammatical: MLU	3.897	0.58	4.433	0.49
Grammatical: DSS	6.824	0.90	7.419	0.96
Grammatical: % correct use of finite verb morphology	81.21	12	93.21	6.2
Standardized assessment of language				
Semantic: CELF Preschool-2 Expressive Vocabulary	10.62	2.15	10.90	2.01
Grammatical: CELF Preschool-2 Word Structure	10.41	2.26	11.14	1.73
Grammatical: CELF Preschool-2 Recalling Sentences	9.17	2.43	10.79	1.90
Attention/Hyperactivity				
BASC-2 Attention problems	47.92	6.25	49.48	5.81
BASC-2 Hyperactivity	48.79	8.21	52.31	7.59
Conners K-CPT 2	61.50	13.96	61.59	17.60
Nonverbal intelligence				
KBIT-2	100.24	13.09	103.83	5.75

Note. NDW = number of different word roots; MLU = mean length of utterance; DSS = developmental sentence score; CELF Preschool-2 = Clinical Evaluation of Language Fundamentals Preschool–Second Edition; BASC-2 = Behavior Assessment System for Children–Second Edition; K-CPT 2 = Kiddie Continuous Performance Test–Second Edition; KBIT-2 = Kaufman Brief Intelligence Test–Second Edition.

Table 3. Analysis of variance: Clinical Evaluation of Language Fundamentals Preschool–Second Edition (CELF Preschool-2).

Variable		<i>df</i>	SS	MS	<i>F</i>	<i>p</i>	<i>d</i>
CELF Preschool-2 Word Structure Scaled	Between groups	1	7.603	7.603	1.880	.176	0.366
	Within groups	56	226.483	4.044			
	Total	57	234.086				
CELF Preschool-2 Expressive Vocabulary Scaled	Between groups	1	1.103	1.103	0.256	.615	0.135
	Within groups	56	241.517	4.313			
	Total	57	242.621				
CELF Preschool-2 Recalling Sentences Scaled	Between groups	1	38.086	38.086	7.991	.007	0.755
	Within groups	56	266.897	4.766			
	Total	57	304.983				
ELI standard	Between groups	1	220.155	220.155	2.320	.133	0.407
	Within groups	56	5314.000	94.893			
	Total	57	5534.155				

Note. *df* = degrees of freedom; SS = sum of squares; MS = mean squares; ELI = Expressive Language Index.

statistically significant group differences were found for MLU, $F(1, 56) = 14.508$, $p < .001$; DSS, $F(1, 52) = 5.452$, $p = .023$; and percentage correct use of finite verb morphology, $F(1, 56) = 22.757$, $p < .001$. The effect sizes for MLU and percentage correct use of finite verb morphology were large ($d = 1.02$ and $d = 1.28$, respectively). The effect size for DSS was medium ($d = 0.65$; see Table 4). It should be noted that four children from the PT group did not have DSS scores because their language samples did not meet the criteria for computing a DSS (i.e., less than 50% of the utterances contained a subject and a verb).

Correlation of Recalling Sentences Scores and Language Sample Variables

A post hoc correlation between Recalling Sentences scores and the language sample analysis variables was conducted. The Recalling Sentences scores were statistically significantly correlated with all of the language sample variables, except for percentage of correct use of finite verb morphology ($\alpha = .05$). The statistically significant

correlations were all of medium size and ranged from .313 to .444.

Attention/Hyperactivity and Nonverbal Intelligence Results

A one-way between-subjects ANOVA was conducted to determine if group differences were present for the non-linguistic variables of attention, hyperactivity, and nonverbal intelligence. Group differences were not observed for either parent report of Hyperactivity, $F(1, 56) = 2.868$, $p = .096$, or Attention Problems, $F(1, 56) = 0.914$, $p = .343$, on the BASC-2 parent rating scales. Effect sizes for these measures were small ($d = 0.45$ and $d = 0.26$). On the performance measure of attention, omission *T* scores obtained from the Conners K-CPT 2 group differences were not observed, $F(1, 55) = 0.000$, $p = .984$, and the effect size was small ($d = 0.0$). Finally, group differences were not observed for scores on the Nonverbal Intelligence subtest of the Kaufman Brief Intelligence Test, $F(1, 56) = 1.825$, $p = .182$. The effect size was small ($d = 0.36$; see Table 5).

Table 4. Analysis of variance: Language sample analysis.

Variable		<i>df</i>	SS	MS	<i>F</i>	<i>p</i>	<i>d</i>
Number of different words	Between groups	1	2271.879	2271.879	7.612	.008	0.754
	Within groups	56	16714.000	298.464			
	Total	57	18985.879				
Semantic analysis score	Between groups	1	492.431	492.431	6.883	.011	0.70
	Within groups	56	4006.483	71.544			
	Total	57	4498.914				
Mean length utterance	Between groups	1	4.164	4.164	14.508	< .001	1.02
	Within groups	56	16.071	0.287			
	Total	57	20.235				
Developmental sentence score	Between groups	1	4.755	4.755	5.452	.023	0.65
	Within groups	52	45.357	0.872			
	Total	53	50.113				
% Correct use of finite verb morphology	Between groups	1	0.209	0.209	22.757	< .001	1.28
	Within groups	56	0.514	0.009			
	Total	57	0.722				

Note. *df* = degrees of freedom; SS = sum of squares; MS = mean squares.

Table 5. Analysis of variance: Nonlinguistic factors.

Variable		<i>df</i>	SS	MS	<i>F</i>	<i>p</i>	<i>d</i>
BASC-2 Hyperactivity	Between groups	1	179.379	179.379	2.868	.096	0.45
	Within groups	56	3502.966	62.553			
	Total	57	3682.345				
BASC-2 Attention problems	Between groups	1	33.379	33.379	0.914	.343	0.26
	Within groups	56	2044.207	36.504			
	Total	57	2077.586				
Conners K-CPT 2	Between groups	1	0.106	0.106	0.000	.984	0.0
	Within groups	55	13934.034	253.346			
	Total	56	13934.140				
KBIT Nonverbal	Between groups	1	186.483	186.483	1.825	.182	0.36
	Within groups	56	5721.448	102.169			
	Total	57	5907.931				

Note. *df* = degrees of freedom; SS = sum of squares; MS = mean squares; BASC-2 = Behavior Assessment System for Children–Second Edition; K-CPT 2 = Kiddie Continuous Performance Test–Second Edition; KBIT = Kaufman Brief Intelligence Test.

Research Question 3

Impact of Nonlinguistic Factors

A group difference for only one standardized assessment measure of language (i.e., Recalling Sentences subtest of the CELF Preschool-2) was identified through ANOVA. In order to determine the role that nonlinguistic factors, including attention, hyperactivity, and nonverbal intelligence, play in standardized assessment outcomes, a series of analyses of covariance (ANCOVAs) were utilized. If group differences were reduced or negated after controlling for these factors, it would support the hypothesis that differences seen between the PT and FT groups on the Recalling Sentences scores could be largely attributed to nonlinguistic factors. The results of the ANCOVA indicated that the parent report of Attention Problems did account for a significant amount of the group difference in Recalling Sentences scores, but the group difference remained after accounting for Attention Problems (see Table 6). The parent report of Hyperactivity did not account for a significant amount of the observed group differences in Recalling Sentences scores. A performance-based measure of attention (i.e., Conners K-CPT) did not account for a significant amount

of the observed group differences in Recalling Sentences scores. Likewise, nonverbal intelligence did not account for a significant amount of the observed group differences in Recalling Sentences scores.

Although previous research has not implicated nonlinguistic domains such as inattention, hyperactivity, or nonverbal intelligence as mediating factors in data obtained from language samples, additional ANCOVAs were conducted to determine if the group differences exhibited on the language sample variables would remain after controlling for these nonlinguistic factors. None of the nonlinguistic skills accounted for a significant amount of the observed group differences on the language sample variables.

Discussion

Although the language deficits of children born preterm have been well documented in prior research, these studies have overly relied on the results of standardized assessment at the exclusion of other language assessment methods, such as language sample analysis. In this study, the children in the PT group performed more poorly than the socioeconomic status- and age-matched group of FT peers on each of the variables obtained from the language samples. This included measures assessing semantic skills (e.g., NDW and the semantic analysis score) as well as measures designed to assess grammatical ability (MLU, DSS, and percentage correct use of finite verb morphology). An examination of the effect sizes associated with the group differences exhibited for the variables obtained from the language sample analysis indicates moderate to large effect sizes, denoting practical significance and statistical significance. Measures of grammatical ability tended to have group differences and effect sizes of larger magnitude than the semantic measures. The one exception to this pattern was the DSS, which had a medium effect size. DSS scores were not calculated for four of the children in the PT group because 50% of the sentences contained in their language sample did not contain a subject and a verb (and were therefore not eligible for the calculation of a DSS). A subjective analysis of these language

Table 6. Analysis of covariance: Behavior Assessment System for Children–Second Edition Attention problems

Variable	Type III SS	<i>df</i>	MS	<i>F</i>	Sig.
Corrected model	56.484 ^a	2	28.242	6.251	.004
Intercept	180.987	1	180.987	40.058	.000
BASC Attention problem	18.398	1	18.398	4.072	.048
Group	44.426	1	44.426	9.833	.003
Error	248.499	55	4.518		
Total	6085.000	58			
Corrected total	304.983	57			

Note. SS = sum of squares; *df* = degrees of freedom; MS = mean squares; Sig. = significance level; BASC = Behavior Assessment System for Children.

^aR squared = .185 (adjusted R squared = .156).

samples showed that these children exhibited substantial grammatical difficulties. The results obtained from the language sample variables indicate consistent difficulty with conversational language skills in children who were born preterm, especially in the area of grammar. In addition, these group differences could not be attributed to nonlinguistic factors such as inattention, hyperactivity, or nonverbal intelligence. This further supports the argument that language sample analysis variables are valid measures of linguistic skill.

Although few studies have investigated the expressive language skills of children born preterm through analysis of language samples, Crosbie et al. (2010) and Mahurin Smith et al. (2014) did not find many significant differences between the PT and FT groups on variables obtained from language sample analysis. In light of these studies, our finding that children born preterm exhibited poorer performance on all of the language sample variables of interest was unexpected. One plausible explanation is that the children in this study were of preschool age, whereas Crosbie et al.'s participants were of school age. Studies of the language profiles of children with language disorders suggest that the nature and manifestation of language deficits may change over time (Sun & Wallach, 2014; Tomblin, Zhang, Buckwalter, & O'Brien, 2003).

Contrary to predictions, this study identified group differences between the PT and FT groups on only one of the variables obtained via standardized assessment, the Recalling Sentences subtest of the CELF Preschool-2. This subtest requires the child to repeat sentences of increasing length and grammatical complexity verbatim. Sentence repetition tasks, such as the tasks on the Recalling Sentences subtest of the CELF Preschool-2, have been cited as a particularly robust method for assessing language impairments in children, with some data suggesting that sentence repetition is the best single test for differentiating between children with and without language impairments (Conti-Ramsden, Botting, & Faragher, 2001). The difficulty exhibited by the children in the PT group with sentence repetition may indicate broad-based difficulties with language, as sentence repetition is a multifaceted task that taps into "virtually all" aspects of language (Klem et al., 2015). Additionally, a subsequent analysis of the role that nonverbal factors played in the Recalling Sentences scores indicated that the group differences in this standardized assessment were still present even after controlling for nonverbal factors. This suggests that the group difference seen on the standardized assessment was not entirely due to group differences in nonlinguistic abilities such as attention.

Given that sentence repetition tasks are a robust diagnostic indicator of language impairment, it was not surprising that the Recalling Sentences scores on the CELF Preschool-2 were significantly correlated with all but one of the language sample analysis variables. Although the expressive language processes and demands involved in recalling a sentence and producing sentences spontaneously in discourse are likely different, the correlation between these variables

suggests that both assessments are useful in the evaluation of expressive language and are associated with each other.

A majority of published studies on language outcomes of children born preterm have consistently found group differences on variables obtained from standardized assessments (Casiro et al., 1990; Foster-Cohen et al., 2007; Magill-Evans et al., 2002; van Noort-van der Spek et al., 2012). Given this finding, the lack of group differences exhibited on the majority of standardized assessment variables was somewhat surprising. Several plausible explanations for this finding exist. First, children born preterm are a heterogeneous group, and the sample characteristics of the PT group vary greatly from study to study. The children who participated in this study may have differed from previous studies of the language outcomes of children born preterm in meaningful ways that influenced group performance. First, the mean GA of the children in this study was 31 completed weeks of gestation. Other studies investigating the language outcomes of children born preterm have utilized samples of children born preterm with lower mean birth GAs. Neurodevelopmental outcomes tend to decrease linearly as GA birth decreases; hence, utilizing a sample with a lower birth GA might have resulted in more robust group differences for the standardized assessment variables. Similarly, developmental outcomes of children born preterm are much poorer for children from low socioeconomic households. In this study, all but one of the mothers had completed some amount of college, and over 65% had completed a bachelor's or graduate degree. In the large Foster-Cohen cohort of preschoolers born preterm, 40% of the mothers had not completed secondary school. Given that rates of language impairment have been shown to vary significantly among different socioeconomic levels, this factor may have skewed the current sample toward more favorable language outcomes.

Taken as a whole, the results of this study offer new insight into the language functioning of children who were born preterm. Many variables related to the neurodevelopmental outcomes of these children, including language, attention, and nonverbal intelligence, were assessed. Out of all of these domains of development, the children in the PT group exhibited significantly poorer outcomes on measures of conversational language skills when assessed through language sample analysis. This indicates that, in children born preterm, language difficulties at the discourse level may still exist even when these children appear to be developing typically when assessed by standardized assessments of global language ability, cognition, and attention.

Clinical Implications

The findings from this study have important clinical implications for practitioners who work with children born preterm. The results of this study strongly support the use of language sample analysis as a tool for measuring the expressive language skills of young children. Deficits in conversational skills may be difficult to assess through the traditional use of standardized assessments. Although weakness in the use of finite verb morphology and sentence complexity

are considered to be hallmark indicators of language impairment in young children, standardized assessments of language may not adequately assess these skills in comparison to data obtained from language samples (Dunn et al., 1996).

Given the importance of language sample analysis as a tool for assessing the conversational language skills of young children, clinicians should not overly rely on the results of standardized assessment when working with this population. Reliance on standardized assessment measures in the diagnosis of language deficits in this population may result in some children not obtaining services to address their expressive language challenges. Rather, language sample analysis in this population seems imperative, particularly in cases where nonlinguistic variables are intact. If standardized assessments are used, clinicians should utilize assessments that have been found to be particularly robust indicators of language impairment, such as sentence repetition tasks. This is further supported by the finding from this study that the CELF Preschool-2 Recalling Sentences scores were significantly correlated with nearly all of the language sample analysis variables.

In order to establish a full picture of the functioning of a child's linguistic system, it is important to consider the results of standardized assessment in conjunction with performance on measures obtained from language sample analysis. This study adds to the growing body of research that underscores the clinical utility of using language sample analysis in the assessment of language development. Language sample analysis is a valid and reliable assessment of language and is able to detect language deficits that are unable to be detected through standardized assessment (Dunn et al., 1996) and remains an invaluable tool.

Study Limitations

Although this study provides important information regarding the unique language profiles of preschoolers born preterm and the role of nonlinguistic factors in the standardized assessment scores, the results of this study should be considered in the context of the limitations of the study. Several issues remain to be clarified through future studies. Due to the characteristics of the sample in this study, the results may not be fully generalizable to other populations of children born preterm. First, the children who participated in this study came primarily from monolingual homes where standard American English was spoken. As American society continues to become more pluralistic, children are increasingly being raised in households where more than one language is spoken or a nonstandard dialect of English is spoken. Further investigation into these populations may reveal interesting findings and improve the generalizability of the results.

Second, because expressive language was the focus of this study, another issue that remains to be clarified is the receptive language abilities of children born preterm. There is some evidence that children who were born preterm have poorer language comprehension than language production, especially for children who have a history of chronic lung

disease (Lewis et al., 2002; Singer et al., 2001). It is possible that nonlinguistic factors such as attention may impact receptive language scores differently than expressive language scores. Future research in this area may help elucidate these complex relationships.

Conclusions

The language skills of children born preterm and full term were assessed via both language sample analysis and standardized, norm-referenced assessment. The children in the PT group performed more poorly than the children in the FT group on all of the measures obtained from the language samples and one measure of language obtained from the standardized assessments. Generally, the children born preterm performed more poorly when language skill was measured through language sample analysis than standardized assessment. Additionally, poor performance on a standardized assessment of sentence repetition could not be explained by nonverbal factors such as nonverbal intelligence, attention, or hyperactivity. We conclude that children born preterm may exhibit language deficits in discourse-level semantic and grammatical skills that are not evident from standardized assessment.

Acknowledgments

This study was supported by the Council of Academic Programs in Communication Sciences and Disorders PhD Scholarship and the ASHA Foundation Student Research Grant in Early Childhood Language Development. The children in the preterm group were recruited from a larger study supported by National Institute on Deafness and Other Communication Disorders Grant R01 DC003311 (awarded to Steven M. Barlow). The authors would like to thank the children and parents who participated in the study. Assistance with data collection and analysis was provided by Cintamy Budiardjo, Veronica Fox, Abby Hill, and Callie Flading.

References

- Aram, D., Morris, R., & Hall, N. (1993). Clinical and research congruence in identifying children with specific language impairment. *Journal of Speech and Hearing Research, 36*(3), 580–591.
- Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. S. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm: A meta-analysis. *Journal of the American Medical Association, 288*(6), 728–737.
- Bos, A. F., Van Braeckel, K., Hitzert, M. M., Tanis, J. C., & Roze, E. (2013). Development of fine motor skills in preterm infants. *Developmental Medicine & Child Neurology, 55*(Suppl. 4), 1–4.
- Casiro, O., Moddemann, D., Stanwick, R., Panikkar-Thiessen, V., Cowan, H., & Cheang, M. (1990). Language development of very low birth weight infants and fullterm controls at 12 months of age. *Early Human Development, 24*(1), 65–77.
- Cole, K., Mills, P., & Dale, P. (1989). Examination of test-retest and split-half reliability for measures derived from language samples of young handicapped children. *Language, Speech, and Hearing Services in Schools, 20*(3), 259–268.

- Conners, K. (2015). *Conners Kiddie Continuous Performance Test—Second Edition Technical Manual*. North Tonawanda, NY: Multi-Health Systems.
- Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *The Journal of Child Psychology and Psychiatry*, 42(6), 741–748.
- Costanza-Smith, A. (2010). The clinical utility of language samples. *SIG 1 Perspectives on Language Learning and Education*, 17, 9–15.
- Crosbie, S., Holm, A., Wandschneider, S., & Hemsley, G. (2010). Narrative skills of children born preterm. *International Journal of Language & Communication Disorders*, 46(1), 83–94.
- Dunn, M., Flax, J., Sliwinski, M., & Aram, D. (1996). The use of spontaneous language measures as criteria for identifying children with SLI: An attempt to reconcile clinical and research incongruence. *Journal of Speech and Hearing Research*, 39(3), 643–654.
- Eisenberg, S. L., Fersko, T. M., & Lundgren, C. (2001). The use of MLU for identifying language impairment in preschool children: A review. *American Journal of Speech-Language Pathology*, 10(4), 323–342.
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories*. San Diego, CA: Singular.
- Foster-Cohen, S. H., Edgin, J. O., Champion, P. R., & Woodward, L. J. (2007). Early delayed language development in very preterm infants: Evidence from the MacArthur-Bates CDI. *Journal of Child Language*, 34(3), 655–675.
- Foster-Cohen, S. H., Friesen, M. D., Champion, P. R., & Woodward, L. J. (2010). High prevalence/low severity delay in preschool children born very preterm. *Journal of Developmental & Behavioral Pediatrics*, 31(8), 658–667.
- Goffman, L., & Leonard, J. (2000). Growth of language skills in preschool children with specific language impairment: Implications for assessment and intervention. *American Journal of Speech-Language Pathology*, 9(2), 151–161.
- Grunau, R. E., Whitfield, M. F., & Davis, C. (2002). Pattern of learning disabilities in children with extremely low birth weight and broadly average intelligence. *Archives of Pediatric & Adolescent Medicine*, 156(6), 615–620.
- Heilmann, J. (2010). Myths and realities of language sample analysis. *SIG 1 Perspectives on Language Learning and Education*, 17(1), 4–8.
- Hughes, D. L., Fey, M. E., & Long, S. H. (1992). Developmental sentence scoring: Still useful after all these years. *Topics in Language Disorders*, 12(2), 1–12.
- Ingram, D. (1974). The acquisition of the English verbal auxiliary and copula in normal and linguistically deviant children. *ASHA Monographs*, 18, 5–14.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test—Second Edition*. San Antonio, TX: The Psychological Corporation.
- Kerr-Wilson, C. O., Mackay, D. F., Smith, G. C. S., & Pell, J. P. (2012). Meta-analysis of the association between preterm delivery and intelligence. *Journal of Public Health*, 34(2), 209–216.
- Klem, M., Melby-Lervåg, M., Hagtvet, B., Halaas Lyster, S., Gustafsson, J., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, 18(1), 146–154.
- Leadholm, B., & Miller, J. (1992). *Language sample analysis: The Wisconsin guide*. Madison, WI: Wisconsin Department of Public Instruction.
- Lee, L. (1974). *Developmental sentence analysis: A grammatical assessment procedure for speech and language clinicians*. Evanston, IL: Northwestern University Press.
- Lewis, B. A., Singer, L. T., Fulton, S., Salvator, A., Short, E. J., Klein, N., & Baley, J. (2002). Speech and language outcomes of children with bronchopulmonary dysplasia. *Journal of Communication Disorders*, 35(5), 393–406.
- Loeb, D. F., Imgrund, C. M., Lee, J., & Barlow, S. M. (2018). Preterm neurodevelopmental outcomes following orosensory entrainment intervention. *Journal of Neonatal Nursing*, 24, 203–207. <https://doi.org/10.1016/j.jnn.2017.11.001>
- Magill-Evans, J., Harrison, M. J., Van der Zalm, J., & Holdgrafer, G. (2002). Cognitive and language development of healthy preterm infants at 10 years of age. *Physical & Occupational Therapy in Pediatrics*, 22(1), 41–56.
- Mahurin Smith, J., DeThorne, L. S., Logan, J. A. R., Channell, R. W., & Petrill, S. A. (2014). Impact of prematurity on language skills at school age. *Journal of Speech, Language, and Hearing Research*, 57(3), 901–916.
- Mahurin-Smith, J., DeThorne, L. S., & Petrill, S. A. (2017). Longitudinal associations across prematurity, attention, and language in school-age children. *Journal of Speech, Language, and Hearing Research*, 60(12), 3601–3608.
- Malvern, D., & Richards, B. (2012). Measures of lexical richness. In C. A. Chapelle (Ed.), *The Encyclopedia of applied linguistics*. <https://doi.org/10.1002/9781405198431.wbeal0755>
- Mangrin, K. S., Horwood, L. J., & Woodward, L. J. (2017). Cognitive development trajectories of very preterm and typically developing children. *Child Development*, 88(1), 282–298.
- Marquardt, T. P., & Gillam, R. B. (1999). Assessment in communication disorders: Some observations on current issues. *Language Testing*, 16(3), 249–269.
- Martin, J. A., Hamilton, B. E., Osterman, M. H. S., Driscoll, A. K., & Matthews, T. J. (2017). Births: Final data for 2015. *National Vital Statistics Reports*, 66(1), 1–69.
- Miller, J. (1981). *Assessing language production in children*. Boston, MA: Allyn & Bacon.
- Miller, J., & Chapman, R. (2007). *SALT: Systematic analysis of language transcripts*. Madison, WI: University of Wisconsin–Madison, Waisman Center, Language Analysis Laboratory.
- Paranka, M. S., Yoder, B. A., & Brehm, W. (1999). Improved outcome of extremely premature infants in the 1990s. *Military Medicine*, 164(8), 568–571.
- Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology*, 5(2), 5–21.
- Polin, R. A., Carlo, W. A., & Committee on Fetus and Newborn. (2014). Surfactant replacement therapy for preterm and term neonates with respiratory distress. *Pediatrics*, 133(1), 156–163.
- Putnick, D. L., Bornstein, M. H., Eryigit-Madzwamuse, S., & Wolke, D. (2017). Long-term stability of language performance in very preterm, moderate-late preterm, and term children. *The Journal of Pediatrics*, 181, 74–79.
- Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior Assessment System for Children—Second Edition*. San Antonio, TX: Pearson.
- Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research*, 39(6), 1239–1257.
- Rice, M. L., & Wexler, K. (2001). *Test of early grammatical impairment*. San Antonio, TX: The Psychological Corporation.
- Sansavini, A., Guarini, A., Justice, L. M., Savini, S., Broccoli, S., Alessandroni, R., & Faldella, G. (2010). Does preterm birth

-
- increase a child's risk for language impairment. *Early Human Development*, 86(12), 765–772.
- Singer, L. T., Siegel, A. C., Lewis, B., Hawkins, S., Yamashita, T., & Baley, J.** (2001). Preschool language outcomes of children with history of bronchopulmonary dysplasia and very low birth weight. *Journal of Developmental & Behavioral Pediatrics*, 22(1), 19–26.
- Sun, L., & Wallach, G. P.** (2014). Language disorders are learning disabilities: Challenges on the divergent and diverse paths to language learning disability. *Topics in Language Disorders*, 34(1), 25–38.
- Tomblin, J. B., Records, N. L., & Zhang, X.** (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research*, 39(6), 1284–1294.
- Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M.** (2003). The stability of primary language impairment four years after kindergarten diagnosis. *Journal of Speech, Language, and Hearing Research*, 46(6), 1283–1296.
- Ukrainetz, T. A., & Bloomquist, C.** (2002). The criterion validity of four vocabulary tests compared with a language sample. *Child Language Teaching and Therapy*, 18(1), 59–78.
- van Noort-van der Spek, I. L., Franken, M. J. P., & Weisglas-Kuperus, N.** (2012). Language functions in preterm-born children: A systematic review and meta-analysis. *Pediatrics*, 129(4), 745–754.
- Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W.** (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech and Hearing Research*, 38(6), 1349–1355.
- Wiig, E., Secord, W., & Semel, E.** (2004). *Clinical Evaluation of Language Fundamentals Preschool–Second Edition*. San Antonio, TX: Harcourt Assessment.
- Zimmerman, E.** (2017). Do infants born very premature and who have very low birth weight catch up with their full term peers in their language abilities by early school age. *Journal of Speech, Language, and Hearing Research*, 61(1), 53–65.