

Research Article

Language Sampling for Preschoolers With Severe Speech Impairments

Cathy Binger,^a Jamie Ragsdale,^a and Aimee Bustos^a

Purpose: The purposes of this investigation were to determine if measures such as mean length of utterance (MLU) and percentage of comprehensible words can be derived reliably from language samples of children with severe speech impairments and if such measures correlate with tools that measure constructs assumed to be related. **Method:** Language samples of 15 preschoolers with severe speech impairments (but receptive language within normal limits) were transcribed independently by 2 transcribers. Nonparametric statistics were used to determine which measures, if any, could be transcribed reliably and to determine if correlations existed between language sample measures and standardized measures of speech, language, and cognition.

Results: Reliable measures were extracted from the majority of the language samples, including MLU in words, mean number of syllables per utterance, and percentage of comprehensible words. Language sample comprehensibility measures were correlated with a single word comprehensibility task. Also, language sample MLUs and mean length of the participants' 3 longest sentences from the MacArthur–Bates Communicative Development Inventory (Fenson et al., 2006) were correlated.

Conclusion: Language sampling, given certain modifications, may be used for some 3-to 5-year-old children with normal receptive language who have severe speech impairments to provide reliable expressive language and comprehensibility information.

In both clinical and research settings, the use of language samples to assess language abilities often is not considered for children with severe speech impairments. For example, in studies focusing on children with language impairment, children with limited intelligibility are often excluded (e.g., Rice, Redmond, & Hoffman, 2006; Rice, Smolik, Rytting, & Blossom, 2010), despite the fact that speech impairment often co-occurs with language impairment. This practice is understandable, as collecting reliable and valid expressive language information for children with severe speech impairments presents significant challenges; spoken language is necessarily mediated by the presence of the speech impairment. However, finding ways to reliably assess spoken language is crucial for these children. Developing methodologies to accurately assess expressive language would allow for greater participation in clinical studies as well as increase the accuracy of expressive language profiles for children with severe speech impairments.

According to Shriberg et al. (2010), children with severe speech impairments may present with a wide variety of disorder typologies, which can be divided into two broad categories—speech delays and motor speech disorders—and then further subdivided into additional categories. Speech delays, also commonly referred to as speech sound disorders (and formerly known as phonological disorders), “typically normalize with treatment” (Shriberg et al., 2010, p. 797). In contrast, motor speech disorders tend to be more severe in nature and may not normalize with treatment. Shriberg et al.’s suggested subcategories for motor speech disorder include childhood apraxia of speech (CAS), dysarthria, and not otherwise specified.

For the purposes of the current study, the term *severe speech impairment* is used to encompass children who have either a speech sound disorder (speech delay) or a motor speech disorder. Although most published speech-language pathology literature pertaining to children with severe speech impairments focuses on particular diagnoses, such as speech sound disorders (e.g., Mortimer & Rvachew, 2010), dysarthria (e.g., Hustad, Gorton, & Lee, 2010; Hustad, Schueler, Schultz, & DuHadway, 2012), or CAS (McNeill & Gillon, 2013), obtaining reliable differential diagnoses for children with severe speech impairments is a complex endeavor with preschoolers presenting the greatest challenge (e.g., Murray,

^aUniversity of New Mexico, Albuquerque
Correspondence to Cathy Binger: cbinger@unm.edu
Editor: Marilyn Nippold
Associate Editor: Lynn Williams
Received July 8, 2015
Revision received October 8, 2015
Accepted December 17, 2015
DOI: 10.1044/2016_AJSLP-15-0100

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

McCabe, & Ballard, 2014; Shriberg et al., 2010). This issue is highlighted, for example, in the recent addition of the “not otherwise specified” category for motor speech disorders (Shriberg et al., 2010) and is further exemplified in recent work questioning the nature of speech impairments in children with autism (Shriberg, Paul, Black, & van Santen, 2011). Further, the investigation of certain constructs and procedures may warrant studying children with severe speech impairments from a broader perspective—that is, including young children in research who have a range of severe speech impairments (i.e., both speech sound and motor speech disorders) as well as those with no differential diagnosis. For example, expressive language test results will be impacted negatively for all children with very poor intelligibility, regardless of the underlying cause, and an examination of this phenomenon may warrant the inclusion of children with a wide array of speech profiles.

Stating precisely what constitutes a severe speech impairment presents its own challenges; using intelligibility or comprehensibility measures is one approach. In the classic approach, word or sentence lists are recited and then transcribed without contextual cues to determine intelligibility (e.g., Yorkston & Beukelman, 1981). In contrast, comprehensibility is defined as “the extent to which a listener understands utterances produced by a speaker in a communication context” (Barefoot, Bochner, Johnson, Ann, & College, 1993, p. 32). With comprehensibility, contextual influences—including linguistic and social contexts—are taken into account, and measures are designed to estimate performance in natural settings (Yorkston, Strand, & Kennedy, 1996). Regardless of the approach, the more severe the speech impairment, the lower the intelligibility or comprehensibility score. Across a range of measures, researchers have largely agreed that a reasonable intelligibility or comprehensibility cutoff for the “severe” (versus a mild or moderate) speech impairment is 50%–60%. As examples, Gordon-Brannan and Hodson (2000) used continuous speech samples; Hustad, Allison, McFadd, and Riehle (2014) used language samples; and Monsen (1981) used a single word intelligibility task, with all considering intelligibility measures below 50%–60% to indicate severely impaired speech. In a similar manner, a commonly used criterion to include children with highly unintelligible speech who require augmentative and alternative communication (AAC) in research studies is having speech less than 50% comprehensible on a single word comprehensibility task (Binger, Kent-Walsh, Berens, Del Campo, & Rivera, 2008; Binger & Light, 2007; King, Binger, & Kent-Walsh, 2015). Thus, despite differences in tasks, an overall pattern for children considered to have severe speech impairments is apparent with a maximal intelligibility (or comprehensibility, depending on the study) of approximately 50%–60%.

All children falling into this category, regardless of the nature of their speech impairment or delay, require accurate assessment of expressive language abilities—a task that is both necessary and complex (Hodson, Scherz, & Strattman, 2002). These children are at risk for having accompanying language impairments, so assessment of this domain is

essential. However, attaining accurate measures of expressive language abilities in the presence of unintelligible (and sometimes extremely limited) speech presents significant challenges.

Language Assessment and Speech Impairments

Accurate assessment of language for children with severe speech impairments is important for several reasons. First, many children with severe speech impairments—including those eventually diagnosed with either speech sound or motor speech disorders—may require intervention not only for speech but also for language (e.g., McNeill & Gillon, 2013; Mortimer & Rvachew, 2010), but language may be overlooked in favor of focusing on the most obvious issue. In addition, and somewhat to the converse, language expectations may be set too low for many of these children—that is, clinicians may underestimate the language potential for a child whose speech is unintelligible, in particular for children with motor speech disorders; that is, the speech impairment may mask linguistic competence, a phenomenon illustrated by studies showing that preschoolers with severe speech impairments can rapidly learn to create multi-symbol messages using graphic symbols even though the ability to do so may not be reflected clearly in their spoken language (Binger, Kent-Walsh, et al., 2008; Binger, Kent-Walsh, Ewing, & Taylor, 2010; Binger, Kent-Walsh, King, Webb, & Buenviaje, in press; Kent-Walsh, Binger, & Buchanan, 2015; King et al., 2015).

Of course, assessing the language skills of children with severe speech impairments is a complex endeavor. It is fortunate that accurate measures of receptive language can be attained by using tests that require no spoken responses, such as the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2006) and the Test for Auditory Comprehension of Language–Third Edition (TACL-3; Carrow-Woolfolk, 1999). However, almost by definition, standardized tests of expressive language—in particular subtests examining semantics and grammar—require children to talk, and scores therefore are impacted negatively by a child’s poor intelligibility. In some cases, even if a test or tool separates receptive and expressive skills, combined scores are still reported (e.g., Vineland Adaptive Behavior Scales; Sparrow, Cicchetti, & Balla, 2005) and therefore can reflect poorly—and inaccurately—on overall language abilities. Despite the fact that these issues have long been discussed in the AAC literature (Beukelman & Mirenda, 2013; Glennen & DeCoste, 1997), language measures requiring the use of spoken language are still reported at times even in large-scale research involving children with severe speech impairments and often with no discussion of how such measures are mediated by the presence of the children’s speech impairments (e.g., Vos et al., 2014).

Using nonstandardized expressive language tools can assist with determining a child’s true expressive language potential. Emerging evidence indicates that dynamic assessment may be used to explore expressive language options

using AAC (Binger, Kent-Walsh, & King, in press; King et al., 2015), and more traditional options, such as language sampling, also may prove viable. In the latter case, transcripts certainly will be impacted by the child's intelligibility. However, it may be possible to extract valuable expressive language information to help guide intervention. For example, clinical experience indicates that intervention for many children with severe speech impairments may focus primarily on speech production and perhaps single-word semantic skills. However, if language sampling analysis indicates that a child's mean length of utterances (MLU) exceeds 1.0, the child is moving beyond the single-word stage, and intervention should support age- and stage-appropriate syntactic and morphological development. Thus, nonstandardized language tools, including the use of AAC and language sampling, may assist with establishing appropriate language expectations for children with severe speech impairments.

Speech and Language Sampling

Speech and language samples have been used for decades to describe the language skills of children with and without language impairments (e.g., Miller & Chapman, 1981). Connected speech samples are considered one of the most socially valid measures of speech and language (Flipsen, 2006a; Kwiatkowski & Shriberg, 1992) and have long been used to assess both phonology and language. With regard to phonology, reliable speech-related measures have been reported in connected speech samples for children with speech delays, at least with measures involving broad transcription (e.g., Barnes et al., 2009; Shriberg, Kwiatkowski, & Hoffman, 1984; Shriberg & Lof, 1991). For example, Barnes et al. (2009) used connected speech samples to examine the phonological accuracy and speech intelligibility of children with Fragile X and Down syndrome. Connected speech samples taken during administration of an autism assessment battery were collected for each participant. All words were transcribed, using one X for each unintelligible syllable, and then transcribed again using narrow phonetic transcription, using the available contextual information to assist with transcription (e.g., repeated viewings of videotaped sessions). These procedures are typical of speech sampling studies, and using contextual information may be particularly important for children whose speech is compromised.

Samples of a child's spoken language also may be used to analyze expressive language skills. Language samples typically are based on language produced during play and have long been used to document the expressive language development of preschoolers with and without language impairments (e.g., Miller, 1981; Miller & Chapman, 1981). Measures commonly reported include mean length of utterance morphemes (MLUm) or in words (MLUw), number of different words, and type-token ratio. Using language sampling measures for children with highly unintelligible speech is less common for obvious reasons; creating accurate transcripts with semantically rich information is challenging when the child's speech is largely unintelligible. However,

techniques for analyzing the language of children with less intelligible speech are emerging in the literature.

Language Samples for Children With Severe Speech Impairments

Multiple researchers have suggested ways to account for unintelligible words and utterances from language samples (e.g., Flipsen, 2006a; van Dijk & van Geert, 2005), which may aid in the transcription of severely impaired speech. For example, Barnes et al. (2009) marked each unintelligible word or syllable in language sampling transcripts with an X. This approach renders it possible to still calculate MLUw, mean number of syllables per word, or mean number of syllables per utterance derived from the sample (Flipsen, 2006a). These types of measures have been shown to be reliable with trained coders for children who would be classified by Shriberg et al. (2010) as having speech delays (e.g., Mortimer & Rvachew, 2010), but children with motor speech disorders or with no reliable diagnoses have received limited attention to date. One exception is a report on the number of syllables vocalized during intervention sessions for six children between ages 2;11 (years;months) and 6;4 with a range of speech impairments (three of whom had or were suspected of having motor speech disorders) and who had comprehensibility of less than 50% on a single word comprehensibility task (Binger, Berens, Kent-Walsh, & Taylor, 2008) with reliable findings reported for this measure.

In addition, Hustad et al. (2014) reported on the use of language sampling with 27 children ages 2;0 to 2;5 with cerebral palsy. Children were placed in one of three clusters, depending on their speech abilities: not talking (44%), emerging talkers (41%), and established talkers (15%). Differential diagnostic information regarding the participants' speech impairments was not reported for these very young children, although it may be expected that most possessed some form of dysarthria. Language samples were based on 10-min play-based interactions between each child and parent. Transcripts were analyzed in two primary ways: (a) for utterances that were completely intelligible, MLUm, number of different words, and number of total words were calculated, and (b) for the entire transcripts (including all vocalizations regardless of intelligibility), both the number of vocal utterances and percentage of intelligible utterances were calculated. Vocal utterances included all types of vocalizations, such as babbling, jargon, unintelligible words, and word approximations. No attempt was made to determine word boundaries or the number of syllables for utterances containing unintelligible vocalizations. Transcript reliability was calculated for 10 of the transcripts for all five key measures with percentage agreement at or above 90% for all measures.

The work of Hustad et al. (2014) represents an important step forward in characterizing the language of children with speech impairments. However, this work would benefit from expansion in several ways: First, it is not currently known if measures such as MLU can be

extracted from the language samples of children with a wider range of speech impairments, including those with CAS or no diagnosis. Second, although Hustad and colleagues reported acceptable levels of reliability for key measures, intelligibility was only examined at the utterance level (i.e., percentage of fully intelligible utterances), and it is not known if reliability can be achieved at the word level (i.e., percentage of intelligible words). Last, all of the participants in Hustad's study were toddlers; establishing language sampling techniques for preschoolers with severe speech impairments—who, due to maturation, are likely to have more speech output—is still needed. If using language sampling with preschoolers with a range of speech impairments is viable, important advances could be made in evaluating the expressive language of these preschoolers. For example, obtaining reliable measures such as MLUw could be helpful in determining whether or not expressive language intervention is warranted in addition to focusing on the child's speech.

Beyond extracting measures from the language sample itself, it would be useful to examine how such measures relate to other clinical and research tools. For example, demonstrating a relationship between the percentages of words understood in a language sampling context with single word comprehensibility tools, such as the Index of Augmented Speech Comprehensibility in Children (I-ASCC; Dowden, 1997), would support the use of this clinical tool for measuring comprehensibility. Although connected language samples may yield more socially valid measures of comprehensibility than a tool such as the I-ASCC (given the full array of both linguistic and situational cues that are present in a language sample), collecting and analyzing such samples is highly time-consuming. Securing measures of comprehensibility are particularly important clinically for children with severe speech impairments as these measures can be used not only to track progress but also to secure funding for AAC devices, and clinicians need efficient tools for this purpose. The I-ASCC is one such measure, which involves recording approximately 30 age-appropriate single word utterances that are later transcribed by either familiar or unfamiliar people. This is a far less time-consuming task than transcribing a 20-min language sample by repeatedly viewing a videotaped play session. Demonstrating a relationship between single word comprehensibility with language sample comprehensibility would lend credibility to the use of the I-ASCC for clinical decision making.

In a similar manner, if reliable language measures such as MLUw can be derived from language samples of children with severe speech impairments, comparing this measure with other known, valid measures of expressive syntax would demonstrate that the MLUw from the transcripts is tapping into a similar construct. For example, for one portion of the Words and Sentences protocol for the MacArthur–Bates Communicative Development Inventory (CDI) parent report instrument (Fenson et al., 2006), parents record their child's three longest sentences. The mean length of these three messages (MLU3) has been used as a

measure of the child's syntactic potential and has been correlated with the MLUw from language samples of children with a wide range of profiles and impairments (Fenson et al., 2006). It would be useful to show if the MLUw derived from language samples of children with severe speech impairments are correlated with parental reports of their children's expressive language abilities, such as MLU3.

Last, the underlying cognitive and linguistic abilities of many children with severe speech impairments, particularly those with motor speech disorders (which may or may not be accompanied by other neuromotor disorders), often are underestimated. For example, teachers' expectations of students tend to decrease for children who appear to have severe disabilities (Cook, 2001), and adults tend to assume that children with speech impairments (as well as those with language impairments) have lower intellectual functioning (e.g., Rice, Hadley, & Alexander, 1993). The repercussions of such assumptions can have dramatic, negative effects on many aspects of a child's life, not the least of which includes educational placement decisions. For example, more than half of children with multiple disabilities in the United States spend more than half of their school days in settings other than general education classrooms (National Center for Educational Statistics, 2013). These kinds of decisions made in the early years of a child's life can have lifelong repercussions for educational, social, and employment outcomes (Lund & Light, 2007). Given the high-stakes consequences of underestimating abilities, understanding the relationship—or lack of relationship—between the severity of a speech impairment and accompanying language and cognitive skills is needed. For children who have severe speech impairments and receptive language within normal limits, demonstrating that no direct link exists between the severity of their speech impairment and their language and cognitive functioning is essential.

Therefore, the overall purposes of the current project are threefold: (a) to determine which aspects of the language samples of children with severe speech impairments can be reliably transcribed; (b) to derive useful linguistic measures on the basis of these samples; and (c) to determine if language sampling measures and comprehensibility are correlated with existing measures of speech, language, and cognition. The following specific research questions were addressed.

Research Question 1

Can point-by-point agreement be obtained for key measures, including word presence, syllable presence, and exact words, from language samples of preschoolers with poor comprehensibility and receptive language within normal limits? We predicted that word presence and syllable presence could be reliably transcribed for some children, but that reliability for exact words would not be established. Given the wide range of possible participants in terms of the nature and severity of the impairment, variability across participants was anticipated.

Research Question 2

Can summary measures, including MLU_w and mean syllables per utterance, be reliably derived from these same language samples? We predicted that reliability would be established for some children for MLU_w and mean syllables per utterance for the reasons stated above.

As we did not anticipate acceptable reliability for exact word productions, we did not attempt to analyze grammatical morphemes or calculate MLU_m. Such analyses likely would have resulted in even poorer reliability, as inflectional morphemes tend to be phonetically minimal and receive little stress. It is notable that Hustad et al.'s (2014) language sampling study of 2-year-olds with cerebral palsy included the use of MLU_m instead of MLU_w. However, the mean MLU_m for the most verbal group, the established talkers, was less than 1.6—that is, even the most advanced children were mainly speaking in one- to two-word messages, at which point a child seldom uses inflectional morphemes (Owens, 2012). Therefore, it may be speculated that, for Hustad et al.'s toddlers, MLU_m and MLU_w measures would essentially have been equivalent (Parker & Brorson, 2005).

Research Question 3

Are the measures derived from the reliably coded language samples correlated with related speech, language, and cognitive assessments? A significant correlation between the percentage of comprehensible words from the language samples and the percentage of comprehensible words on the I-ASCC (Dowden, 1997) was predicted, demonstrating that both measures tap into a similar construct. For the same reason, a significant correlation between the MLU_w from the language samples and MLU₃ from the CDI (Fenson et al., 2006) also was anticipated; lower scores for the MLU_w were expected on the basis of findings with other populations (Fenson et al., 2006). We anticipated no significant correlations between the children's language sample comprehensibility and scores on the TACL-3 (Carrow-Woolfolk 1999), PPVT-4 (Dunn & Dunn, 2006) and Leiter International Performance Scale–Revised (LIPS-R; Roid & Miller, 1997) as these tests require no spoken language and should not be affected by comprehensibility for these participants with intact receptive language.

Method

Participants

Fifteen 3- to 5-year-old participants were included in the current investigation (see Figure 1 for a flowchart of participant inclusion) with all recruited through the University of New Mexico's Speech and Hearing Clinic or local school districts. The University of New Mexico is located in a culturally diverse metropolitan area. Major ethnic and racial populations in the local county include Latinos (49%), Anglo/European Americans (41%), and Native Americans (6%; U.S. Census Bureau, 2014). Study

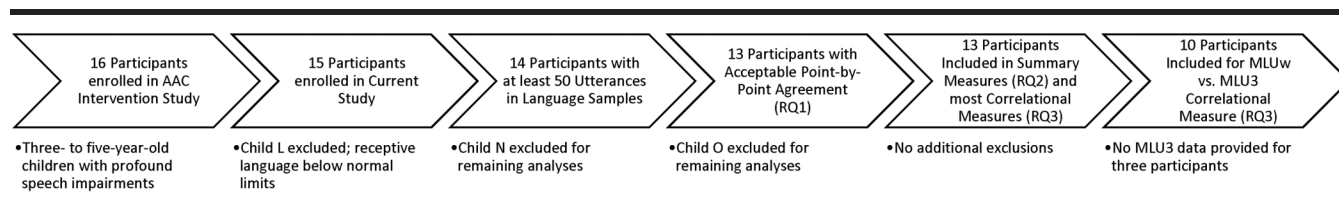
participants reflected this racial and ethnic diversity and included eight children who were White Hispanic, six who were White non-Hispanic, and one Asian Indian/White non-Hispanic child. The primary language for all participants was English, with two participants being raised in bilingual or trilingual environments: Child A was exposed to Hindi and Tamil with an estimated 95% exposure to English and 5% exposure to Hindi and Tamil between ages 2 and 5. Child C was exposed to Spanish but “mostly heard English,” according to her mother; English was her mother's first language and her father's second (fully fluent) language. Second language exposure, then, was unlikely to interfere with the results of this study for these two 5-year-olds.

All participants in the current investigation were included in a broader study designed to investigate the effects of an AAC intervention on the expressive language skills of preschoolers with severe speech impairments. The primary focus of this broader investigation was on AAC intervention, and the differential diagnoses of the participants' speech impairments was not a primary concern. Therefore, comprehensive speech assessments were not conducted. Participants were required to meet the following criteria: (a) age 3;0 to 5;11 at the onset of the investigation; (b) English spoken as the primary language; (c) presence of a severe speech impairment, defined as less than 50% comprehensible language in the “no context” condition of the I-ASCC (Dowden, 1997); (d) standard score no more than 1.5 *SD* below the mean on the TACL-3; (e) expressive vocabulary of at least 25 words/symbols on the CDI (Fenson et al., 2006) via any communication mode (speech, sign, aided AAC); (f) parental report of functional vision for participating in study activities and adequate hearing as determined by a pure-tone hearing screening; and (g) no diagnosis of autism spectrum disorder (see Table 1). Graduate and undergraduate students majoring in speech and hearing sciences who were unfamiliar with the participants judged comprehensibility for the I-ASCC; a different listener was used to score each sample to eliminate task familiarity influences.

One child (Child L) was included in the AAC intervention study but is not included in the current study as his receptive language was not within normal limits. Child H's diagnosis of autism was not known at the onset of the investigation. As this participant demonstrated appropriate social skills throughout the investigation, responded positively to feedback from study administrators, readily completed all investigation tasks, and earned receptive language and IQ scores within normal limits, his data are included.

Additional measures included the following: (a) PPVT-4 (Dunn & Dunn, 2006), a test of receptive vocabulary; (b) LIPS-R (Roid & Miller, 1997), a test of nonverbal intelligence; and (c) Vineland Adaptive Behavior Scales (Sparrow et al., 2005), a parent interview measuring functional adaptive behaviors across various domains (see Table 2). Language sampling and standardized testing were completed for each child directly after consent was attained. Language

Figure 1. Flowchart of participant inclusion.



sampling was completed within one session for all children, and standardized testing was completed across multiple sessions.

With regard to disability type, four children were diagnosed with motor speech disorders, eight with speech sound disorders (several of whom entered the study with inaccurate diagnoses of suspected CAS), and three were unknown. Five participants (Children A, F, H, N, and O) received differential diagnostic evaluations at the University of New Mexico Speech and Hearing Clinic during or after the study, and their diagnoses are considered accurate; diagnoses for the remaining children are based on school reports and observational data and are considered less reliable, given that common assessment practice in New Mexico public schools is to establish the need for special education services without determining differential diagnoses.

Setting and Experimenters

The language samples were collected by the first author (a certified speech-language pathologist) and three speech-language pathology graduate students under the first author's supervision. The first author trained the graduate students prior to administering the language samples and provided coaching and supervision as needed during language sampling sessions. All sessions were conducted at the University of New Mexico Speech-Language and Hearing Clinic in a therapy room.

Language Sampling Session Procedures

Each child participated in one language sampling session that lasted approximately 20 min and was video

Table 1. Participant characteristics.

Child	Sex/CA	Disability	CDI	I-ASCC		Lang Samp Comp
				No context	With context	
A	F/5;10	Motor speech disorder—dysarthria (ataxic) secondary to cerebral palsy	657	13%	53%	75%
B	M/4;11	Speech delay, history of TBI; microdeletion of 7q11.22 ^a	115	0%	3%	52%
C	F/5;1	Speech delay	514	16%	55%	68%
D	M/5;9	Speech impairment—unknown	> 86 ^b	35%	68%	89%
E	M/3;10	Speech delay	745	21%	51%	87%
F	F/4;4	Motor speech disorder—not otherwise specified secondary to Bainbridge-Ropers syndrome	57	0%	6%	33%
G	F/4;8	Speech delay—initially diagnosed with CAS and later changed to speech delay	601	10%	42%	96%
H	M/5;0	Autism, motor speech disorder—initially CAS but later changed to ataxic dysarthria	109	0%	3%	32%
I	M/4;1	Speech impairment—unknown	601	6%	38%	65%
J	M/4;2	Speech delay, history of tongue tie (frenulum cut 1 year prior to onset of study)	323	0%	13%	63%
K	M/4;3	Speech delay	419	3%	12%	72%
M	M/4;3	Speech delay, history of drug exposure in utero, current deficits are speech only	628	0%	26%	58%
N	F/4;9	Motor speech disorder—dysarthria secondary to cerebral palsy	547	3%	16%	—
O	M/3;5	Speech delay	83	6%	23%	—
P	F/3;3	Speech delay	45	0%	13%	81%

Note. CA = chronological age in years;months; CDI = MacArthur–Bates Communicative Development Inventory (Fenson et al., 2006); I-ASCC = Index of Augmented Speech Comprehensibility in Children (Dowden, 1997); Lang samp comp = Percentage of comprehensible words from language sample (mean of the two transcribers); TBI = traumatic brain injury; CAS = childhood apraxia of speech.

^aThis deletion has been associated with autism, but data are incomplete in the research at this time. Child B received diagnostic testing and did not have autism. ^bThe CDI was not completed for Child D. This is the number of different words used in a 20-min language sample taken at the beginning of the study and is a gross underestimate of his expressive vocabulary.

Table 2. Language and cognitive test results and language sampling measures.

Child	PPVT-4		TACL-3		Vineland-Comm		LIPS-R: Full IQ	
	Standard score	Percentile	Standard score	Percentile	Standard score	Percentile	Standard score	Percentile
A	87	19	91	27	100	50	94	34
B	99	47	111	77	87	19	101	53
C	88	21	87	19	83	13	108	70
D	109	73	111	77	95	37	117	87
E	113	81	128	97	104	61	122	93
F	79	8	94	35	65	1	79	8
G	83	13	83	13	91	27	84	14
H	101	53	96	39	81	10	108	70
I	95	37	89	23	69	2	101	53
J	121	92	115	84	106	66	113	81
K	95	37	104	61	83	13	116	86
M	98	45	91	27	95	37	102	55
N	103	58	91	27	87	19	101	53
O	124	95	111	77	85	16	143	> 99
P	92	30	106	65	87	19	114	82

Note. PPVT-4 = Peabody Picture Vocabulary Test—Fourth Edition (Dunn & Dunn, 2006); TACL-3 = Test of Auditory Comprehension of Language—Third Edition, Total Score (Carrow-Woolfolk, 1999); Vineland-Comm = Vineland Adaptive Behavior Scales—Communication subtest (Sparrow et al., 2005); LIPS-R = Leiter International Performance Scale—Revised (Roid & Miller, 1997)

recorded in its entirety. Sony Handicam Digital HD Video Camera Recorders were used to record all sessions. Video recorders were placed on tripods in the same room in which the sampling took place. The camera was moved as needed, typically by a second examiner, to maximize the view of the child's face throughout the session. Additional microphones were not used to collect language samples as the sound quality from the cameras was sufficient for transcription. During each sampling session, the examiner and child were seated on the floor or at a table, depending on the child's needs and preferences. Standard interactive play materials, such as toy picnic items, dolls, and vehicles, were used during language sampling. The specific toys used in each session were adjusted on the basis of each child's interests with individual items made available throughout the sessions as deemed appropriate by the examiner. Typical, recommended language sampling techniques were followed, such as avoiding the use of yes/no questions and other questions that yield one-word responses and allowing the child adequate time to initiate and respond (Miller, 1981).

Transcription Procedures

Two transcribers, who were not involved in collecting language samples, separately completed an orthographic transcription of each language sample for each of the 15 children. The first author trained these two upper-level undergraduate students in speech-language pathology (including the second author) on transcription procedures. Before transcribing samples in the current study, the transcribers independently completed a transcript of a language sample of a typically developing preschooler and were required to obtain point-by-point reliability of at least 90% for both child and adult utterances compared with the transcript of an experienced transcriber as well as with each other. Both transcribers met these criteria.

Transcripts were created on Excel spreadsheets with spreadsheet columns including the following labels: counter number, adult utterances, adult actions, child utterances, and child actions. Counter numbers were recorded approximately every 30 s to enhance the ease of locating transcript position. Adult utterances included all messages spoken by the examiner. Adult and child actions were recorded as needed to clarify session activities (e.g., [Child hands hot-dog to Examiner]). Instructions for recording the children's utterances included the following: (a) transcribe child's messages word for word to the best of your ability; (b) for any word that is transcribed as comprehensible, at least one phoneme should approximate the phonemes in the transcribed word; (c) for noncomprehensible words, use an X to indicate each syllable that cannot be determined even with contextual cues; (d) indicate word boundaries for unintelligible words with spaces—for example, two single-syllable words are transcribed as X [space] X, and one two-syllable word is transcribed as XX; (e) indicate utterance boundaries by placing new utterances on a separate line—consecutive words spoken with less than a 3-s pause in between them are considered one utterance; (f) transcribe as a vocalization [voc] if child produces a general sound such as “mmm” that does not phonetically resemble a contextually appropriate word. More detailed operational definitions were provided when required.

Transcribers were permitted to view all sections of the videos an unlimited number of times, and transcribers were not discouraged from “glossing”—that is, transcribing words that the examiner repeated during the language samples when the children's speech was difficult to understand—a procedure that has been shown to increase the number of words that can reliably be transcribed (Flipsen, 2006b). The frequency of glossed words was not tracked. Collecting 50–100 utterances is recommended when doing language samples with preschoolers, and although 100 utterances

is preferable, a 50-utterance sample will provide about 80% of the same information that 100 utterances would provide (Paul & Norbury, 2012). The first 100 utterances from each 20-min language sample were analyzed for this investigation. If 100 utterances were not available within the sample, all available utterances were used (50 utterances minimum).

Data Analyses

Global Measures

Both transcribers independently completed a transcript for each child. Data on the following global measures were calculated for each transcript: total number of comprehensible words; total number of words (i.e., total number of comprehensible and noncomprehensible words); total number of syllables (i.e., total number of comprehensible and noncomprehensible syllables); and total number of different words. These global measures were used for the summary measure calculations (described below).

Research Question 1: Point-by-Point Agreement

A range of reliability measures was calculated for each pair of transcripts (i.e., both transcripts for Child A, both for Child B, and so on). Two speech-language pathology students masked to the purposes of the study completed these analyses. For two words to agree on the word presence measure, both transcribers had to transcribe a word as occurring at the same point in time on the recording. Credit was given for this measure regardless of whether the words were the same (e.g., *plate* vs. *plate*), different (*plate* vs. *plane*), or unintelligible (*plate* vs. X). In addition, word boundaries were required to be the same for this measure; for example, if one transcriber recorded XX (one two-syllable word) and the other transcribed X X (two one-syllable words), the transcribers were considered to have agreed on the presence of one word but disagreed on the presence of the second. In this example, then, the transcribers would only achieve 50% reliability for word agreement. For credit to be awarded for syllable presence, both transcribers had to indicate the presence of syllables at the same point in time whether these syllables were the same (*airplane* vs. *airplane*), different (*airplane* vs. *apple*), or unintelligible (*airplane* vs. XX). For this measure, identical word boundaries were not required (XX vs. X X counted as agreement for the presence of two syllables). The last point-by-point measure, exact word agreement, is more commonly used to determine language sampling reliability. To receive credit, word roots had to be the same for agreement to be awarded (*plate* vs. *plate* received credit; *plate* vs. *plane* and *plate* vs. X did not). Unintelligible words, exclamations, and fillers (such as *um*) were excluded from the exact word agreement analysis when present on both transcripts (e.g., *um* vs. *mm*, *um* vs. X, or X vs. X).

Cohen's kappa (Viera & Garrett, 2005) was used to calculate interobserver agreement for word presence, syllable presence, and exact word agreement. According to these authors, levels of agreement for the kappa statistic are as follows: < 0 = *less than chance agreement*; 0.01–0.20 = *slight*

agreement; 0.21–0.40 = *fair agreement*; 0.41–0.60 = *moderate agreement*; 0.61–0.80 = *substantial agreement*; 0.81–0.99 = *almost perfect agreement*. We anticipated substantial to almost perfect agreement for the majority of the transcripts for word and syllable presence but less acceptable levels for exact word agreement. Kappa scores of at least 0.61 (substantial agreement) were required for calculating the summary measures (below).

Research Question 2: Summary Measures

Summary measures derived from the global and point-by-point measures were calculated for each transcript and included the following: (a) MLUw (total number of words—both comprehensible and noncomprehensible—divided by the total number of utterances); (b) average number of syllables per message (total number of syllables—both comprehensible and noncomprehensible—divided by the total number of utterances); and (c) percentage of comprehensible words (number of comprehensible words divided by the total number of words).

Spearman's rho was used to determine if the summary measures were reliable with the expectation of a significant correlation between the two transcribers for all summary measures: MLUw, average number of syllables per message, and percentage comprehensibility. Spearman's rho, a nonparametric statistic, uses rank ordering to determine reliability—in this case, the relative rank orders of the transcribers' data for each of these three measures. A significant finding for this statistic would indicate that the two transcribers agreed on the relative ordering of the summary measures (e.g., which children were the most vs. the least comprehensible). As the rho statistic is based on rank orderings only, data may be highly correlated but still differ in terms of their absolute values; therefore, the Wilcoxon signed-ranks test, essentially a nonparametric version of a *t* test, was used to determine absolute differences on the summary measures between transcript pairs.

Research Question 3: Correlations

Correlations between the language sample measures and other language and cognitive measures were calculated for the third research question. For all language sample measures, the mean scores of Coders 1 and 2 were used. Spearman's rho was used for the following comparisons: (a) percentage of comprehensible words from the language samples versus I-ASCC scores (with and without context conditions); (b) MLUw from the language samples versus MLU3 from the CDI (Fenson et al., 2006); and (c) percentage of comprehensible words from the language samples versus the TACL-3 (Carrow-Woolfolk, 1999), PPVT-4 (Dunn & Dunn, 2006), and LIPS-R (ROID & Miller, 1997). As discussed above, Spearman's rho uses comparisons of rank ordering to analyze correlations; in these cases, comparisons were made between transcription data and test scores. Again, mean scores for the two transcribers were used for this analysis.

Results

Transcripts Included and Excluded

Twenty-minute language samples were collected and transcribed for all 15 children. Out of 15 language samples, 11 contained at least 100 utterances. For these samples, the first 100 utterances were used for all calculations below; all utterances were used for samples with fewer than 100 utterances. Samples with fewer than 100 utterances included Children C (93 utterances), E (59), H (88), J (83), and N (17). Data for Child N were excluded from the remaining analyses due to the low number of utterances (see Figure 1). Child N had cerebral palsy and was the most motorically impaired child in the study. She was unable to walk independently and had slow, labored, dysarthric speech. Although she frequently attempted to talk during the language sample, nearly all of her speech attempts were characterized as “vocalizations” by both transcribers, not as words, and therefore were not counted as utterances. Data for the remaining 14 participants were used in the analyses below unless otherwise noted.

Global Measures

Data on global measures revealed that the participants demonstrated a range of expressive language abilities with a high degree of variability. Using the means for each pair of transcripts for each child, findings were as follows: Total number of comprehensible words mean was 127 (range = 37–226, $SD = 69$), total number of words mean was 188 (range = 102–281, $SD = 63$), total number of syllables mean was 216 (range = 115–311, $SD = 74$), and total number of different words mean was 43 (range = 3–86, $SD = 26$).

Research Question 1: Point-by-Point Agreement

For the current study, kappa scores falling within the substantial or almost perfect agreement range were deemed acceptable (0.61 or above; Viera & Garrett, 2005). For word and syllable presence, 13 of 14 transcript pairs met this criterion. For word presence, eight of 14 transcripts fell within the almost perfect range and five within the substantial range ($M = 0.80$). For syllable presence, 11 of 13 were almost perfect and two were substantial ($M = 0.82$). Exact word agreement scores, as expected, were lower with a mean $\kappa = 0.61$ (range = 0.33–0.81). Only one score was within the almost perfect range for exact word agreement (Child G), and five were in the substantial agreement range (see Table 3).

Child O's agreement measures were notably lower than those for all other participants (0.34 for word and syllable presence, 0.33 for exact word agreement). Further analysis revealed that, in many instances, Child O's speech was interpreted by one transcriber as a word (mostly noncomprehensible and therefore transcribed as X) with the other transcriber interpreting these same instances as nonword vocalizations (transcribed as [voc]). Each of these cases was scored as a disagreement, resulting in poor overall agreement. Due to low inter-rater agreement, Child O's data were excluded for the remaining analyses (see Figure 1).

Table 3. Cohen's kappa scores for word presence, syllable presence, and exact word agreement.

Child	Word presence	Syllable presence	Exact word agreement
A	0.84	0.84	0.60
B	0.72	0.75	0.64
C	0.90	0.92	0.58
D	0.90	0.91	0.71
E	0.94	0.96	0.71
F	0.81	0.84	0.64
G	0.91	0.91	0.81
H	0.80	0.83	0.59
I	0.82	0.83	0.57
J	0.80	0.92	0.57
K	0.91	0.92	0.59
M	0.79	0.82	0.52
O	0.34	0.34	0.33
P	0.68	0.68	0.73
Mean	0.80	0.82	0.61

Research Question 2: Summary Measures

Thirteen transcripts were included in the summary measures analyses (all except for Children N and O). Spearman's rho correlations to compare the two transcribers were significant for all three summary measures: MLUw, mean syllables per utterance, and comprehensibility (see Table 4). That is, there was a strong association between the rank orders for the two sets of transcripts for these measures—and for MLUw and mean syllables per utterance, virtually identical.

Wilcoxon signed-ranks test results were significant for MLUw and percentage of comprehensible words but not for mean syllables per utterance; that is, there was a significant difference in the absolute scores between the two transcribers for MLUw and percentage of comprehensible words, but the mean number of syllables per utterance were statistically similar. Further analysis revealed that the values for Transcriber 1 were slightly but consistently lower than for Transcriber 2, which explains the difference between the Spearman's rho versus the Wilcoxon results; that is, the rank ordering for Spearman's rho allows for a difference in scores as long as those differences are consistent, whereas the Wilcoxon test examines differences in the scores themselves. The mean values for the summary measures for Transcriber 1 versus Transcriber 2, respectively, were as follows: MLUw was 2.2 versus 2.1; mean syllables per message was 2.5 versus 2.4; and comprehensibility was 61.4% versus 67.5%. MLUw and mean syllables per utterance for each participant are depicted in Figure 2.

Research Question 3: Correlations

Thirteen transcripts were included in the I-ASCC analyses (all except for Children N and O). A significant correlation was found between comprehensibility from the language samples versus the I-ASCC both with and without contextual cues (see Table 5), indicating that language

Table 4. Results for Research Question 2: Reliability of summary measures.

Measures	Spearman's rho		Wilcoxon <i>W</i>	
	<i>R/R</i> ²	<i>p</i> values	<i>W</i> value/critical value	<i>p</i> values
MLUw	.98/.96	< .001*	0.0/8	Significant
Mean syllables per utterance	.99/.98	< .001*	10.5/10	Not significant
Percentage of comprehensible words	.80/.64	< .001*	13.0/17	Significant

Note. MLUw = mean length of utterances in words.

sample comprehensibility and I-ASCC scores measure a similar construct.

MLUw was compared with the mean of the three longest utterances on the CDI (see Figure 2). Three families failed to fill out this portion of the CDI, and Children N and O were excluded, so calculations were based on the remaining 10 participants. Children lacking an MLU3 score should not be viewed as failing to speak in multiword utterances; rather, their families simply neglected to complete this portion of the CDI. As expected, MLUw scores were lower for nearly all participants (exception: Child F with MLUw = 1.2 and a MLU3 = 1.0). Overall means were 2.1 for MLUw (range = 1.2–2.9) and 5.2 for the CDI measure (range = 1.0–12.3). A significant correlation was found between MLUw and MLU3 (see Table 5). As anticipated, PPVT-4, TACL-3, and LIPS-R scores were not significantly correlated.

Discussion

Viability of the Language Samples

Of the 15 participants, language samples of at least 50 utterances were collected for all but one child, indicating that it is possible to collect language samples for children with a range of severe speech impairments—even children with extremely low single word comprehensibility. The exception was Child N, who had cerebral palsy and was

the most motorically impaired participant. As noted in the results, most of her speech attempts were coded as vocalizations, not as utterances. Two possible solutions to ensure a sufficient number of utterances collected for children with similar profiles are to plan for longer or repeated language samples or to elicit connected speech samples using more structured sentence elicitation tasks (Hustad et al., 2012). Another approach, which has been used for analyzing the language samples of 2-year-olds with cerebral palsy (Hustad et al., 2014), is to count general vocalizations (such as babbling and jargon) as utterances and then calculate the number of vocal utterances and the percentage of intelligible utterances for each child. Although the latter measure may be used to approximate comprehensibility, neither approach measures syntax—a construct typically of primary interest with preschool language samples. A quite different but complementary approach that may be used to examine syntax for children with limited word productions is to explore expressive syntax by using aided AAC. For example, nearly all children in the current investigation (including Child N) learned to produce two- to three-symbol rule-based messages (e.g., agent-action-object sentences) during a relatively brief dynamic assessment task in which children used a communication app on an iPad (Binger, Kent-Walsh, & King, in press; King et al., 2015). Thus, alternative communication modes should be considered when assessing expressive language potential for children with severe speech

Figure 2. Mean length of utterances in words and mean syllables per utterance from the language transcripts and mean length of the three longest utterances from the MacArthur–Bates Communicative Development Inventories for transcripts with acceptable inter-rater reliability. MLUw = mean length of utterance from language samples; Mean syll = mean number of syllables per utterance from language samples; MLU3 = mean of three longest messages recorded on the MacArthur–Bates Communicative Development Inventories (Fenson et al., 2006).

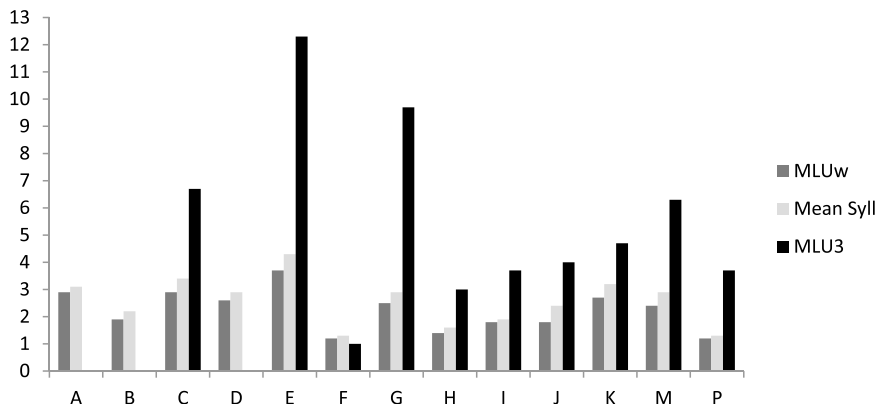


Table 5. Results for Research Question 3: Comparisons of key language sampling measures versus various assessment tools.

Comparison	Percentage comprehensible comparisons (Spearman's rho)		Outcome
	<i>R/R</i> ²	<i>p</i> values	
LSC versus I-ASCC with context	.7800/.6100	< .01*	Significant
LSC versus I-ASCC without context	.7300/.5300	< .01*	Significant
MLUw versus MLU3	.8300/.6900	< .01*	Significant
LSC versus PPVT-4	.0800/.0064	.79	Not significant
LSC versus TACL-3	.0000/.0000	1.00	Not significant
LSC versus LIPS-R	.3300/.1100	.27	Not significant

Note. LSC = language sample comprehensibility; I-ASCC = Index of Augmented Speech Comprehensibility in Children; MLUw = mean length of utterance in words on the language samples, including unintelligible words; MLU3 = mean length of the three longest messages from the MacArthur–Bates Communicative Development Inventory; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; TACL-3 = Test of Auditory Comprehension of Language–Third Edition; LIPS-R = Leiter International Performance Scale–Revised.

**p* ≤ .05.

impairment, particularly for those with the most limited speech output.

For the 14 transcripts that contained at least 50 utterances (all but Child N), global measures were extracted and included the total number of words, comprehensible words, syllables, and different words. These measures varied widely across participants; for example, the range for the number of comprehensible words across transcripts was 37 to 226. These results are not surprising, given the diverse profiles of the participants. These measures were used to calculate word presence, syllable presence, and exact word agreement for each pair of transcripts. Acceptable levels of exact word agreement (e.g., where/where vs. where/here) – that is, kappa scores of at least 0.61 – were only achieved for 5 out of the 14 transcripts. Exact word agreement typically is a key measure when calculating interrater reliability of language samples, and the findings here illustrate why children with severe speech impairments so often are excluded both clinically and in research from language sampling analysis: achieving interrater reliability for exact words may not be possible for many of these children.

However, the fact that acceptable levels of agreement were reached for both word and syllable presence for the majority of the participants is encouraging; that is, substantial or almost perfect agreement was reached across the two transcripts for 13 out of 14 participants for word and syllable presence with the majority falling in the almost perfect agreement range. This finding indicates that these two language sampling measures can be used reliably with at least some—and perhaps, depending on the population studied, most—young children with severe speech impairments. These highly useful numbers can be used, in turn, to calculate MLUw and mean syllables per utterance, thereby providing broad indicators of a child's current level of expressive syntax via spoken language.

The reliability of only one child fell below the substantial agreement indicator for word presence and syllable presence: Child O. A post hoc analysis of Child O's transcripts indicated that, in many instances, one transcriber indicated

the presence of unintelligible words (transcribed as X) when the other indicated the presence of vocalizations (transcribed as [voc]); the latter did not count as words. As discussed above, Hustad et al. (2014) avoided this issue by not differentiating between vocalizations and unintelligible words and basing MLU only on intelligible words. However, in the current study, one of the major purposes was to track the presence of all perceived words (whether intelligible or not) and then derive MLUw from those words, so Hustad et al.'s approach was not viable in the current investigation. As an alternative, tightening operational definitions may have helped in the current study. The operational definitions worked well for differentiating comprehensible versus noncomprehensible words; that is, comprehensible words had to contain at least one phoneme similar to the transcribed word, but noncomprehensible words did not. However, the definitions did not adequately differentiate between noncomprehensible words versus vocalizations. In the latter case, the only direction given was to “transcribe general sounds such as *mmm* as vocalizations (i.e., [voc] in the transcript), not as Xs or as comprehensible words.” Refining this to include the use of contextual cues when differentiating between vocalizations and words might have been of assistance; for example, one indicator of a child using a word instead of a vocalization would be if the child looked at or pointed to something specific while speaking. Continuing to operationalize transcript rules for this population, then, is warranted.

Syntactic and Comprehensibility Measures From Transcripts

With regard to the summary measures for Research Question 2—that is, MLUw, mean number of syllables per message, and comprehensibility percentages—statistical testing revealed a significant level of agreement with all three measures using Spearman's rho calculations; that is, relative rankings for the two transcribers were similar with rankings nearly identical for MLUw and number of syllables per

message. These findings provide an initial indication of the viability of using language sampling methodologies to assess the expressive language of children with severe speech impairments.

It must be noted, however, that differences were apparent across the two transcribers for some measures; that is, Wilcoxon signed-ranks test results indicated a significant difference in the absolute scores for MLUw and comprehensibility. The high correlations with Spearman's rho rankings indicate that these differences must have been consistent; that is, one transcriber consistently gave the children slightly more credit for producing words than the other transcriber. The data indicate that, at least for these two transcribers, these differences may not be clinically significant. Across the two transcribers, the mean MLUw and mean syllables per message only differed by 0.1 (i.e., 2.2 vs. 2.1 and 2.5 vs. 2.4, respectively).

Comprehensibility Correlations

In Research Question 3, the relationship between various language transcript measures and other constructs was examined. First, the percentage of comprehensible words from the language samples was compared with the single word comprehensibility scores from the I-ASCC, with unfamiliar listeners providing I-ASCC ratings. Results for the Spearman's rho analysis were significant; that is, children with lower percentage comprehensibility on the language samples also tended to exhibit lower comprehensibility scores on the I-ASCC, indicating that these two tools may be tapping into the same underlying construct. It was not surprising that a stronger correlation was noted for the with-context condition of the I-ASCC than the without-context condition; both language sampling and the with-context condition of the I-ASCC provide listeners with contextual information to support understanding of the child's language. From a clinical perspective, the findings lend support for using the I-ASCC as an efficient way to measure of comprehensibility for children with severe speech impairments.

A related and somewhat unexpected finding is the stark difference in absolute values between the percentage of comprehensible words from the I-ASCC compared with the transcripts. Even though these measures were correlated, the absolute values varied dramatically for some children. The difference in the means for the with-context scores on the I-ASCC versus language sample comprehensibility (again excluding Children N and O) was 41.5% (i.e., 25.5% vs. 67%, respectively). Scores were expected to increase on the highly contextualized language samples, but not so dramatically. The data set is too small, too varied (in terms of typology), and too uncertain (in terms of diagnoses) to draw any firm conclusions. One trend of interest, however, is the fact that the three children with known motor speech disorders had some of the smallest differences (i.e., less than 30% for Children A, F, and H). In other words, children with speech sound disorders (as opposed to motor speech disorders) may demonstrate more variability across various comprehensibility measures than those with motor

speech disorders; this is not unexpected given differences in the nature of these impairments (Shriberg et al., 2010). Further research is needed to substantiate this hypothesis.

Syntactic Correlations

Both MLUw (from the language samples) and MLU3 (the mean of the three longest messages indicated by parent report on the CDI) are measures of syntactic complexity. As predicted, a significant correlation was found between these two measures. The proportional difference between MLUw and MLU3 was roughly equivalent to the differences found in other studies that examined this relationship; that is, previous studies have reported that MLU3 tends to be approximately 2.5 times the child's MLUw (for a summary of these findings, see Fenson et al., 2006). Given the relative ease of collecting MLU3 and the inherent difficulties involved in securing other valid and efficient measures of expressive syntax for these children, further research to substantiate the use of MLU3 as an estimate of syntactic potential for children with severe speech impairments is justified.

However, caution is certainly warranted, particularly for children with motor speech disorders. First, it must be stressed that the norms reported for the CDI are not applicable for this population. Second, children with motor speech disorders (CAS, dysarthria, or nonspecified) may produce very little speech (e.g., Hustad et al., 2014) even compared with most of the children in the current study, but it does not follow that such children lack expressive language potential or abilities. For example, Child N—one of four participants with a motor speech disorder—was excluded from the analyses as she produced so few messages in the language sample. The vast majority of her transcribed speech was characterized as general vocalizations, and all but three of her actual utterances were single-syllable words. However, this child evidenced use of rule-based three-word messages using aided AAC during a dynamic assessment task and in a subsequent AAC intervention (e.g., "Lion chase sheep," "Sheep is happy"), and her MLU3 on the CDI was 3.5. Her ability to create such multiword messages was not apparent in her language sample. In a similar manner, Child O, whose summary measures were not analyzed due to poor inter-rater reliability, had an MLU3 of 3.7. These findings indicate the need to develop alternative means of assessing the expressive language skills of children for whom even the modified language sampling used in the current study is not a viable option. To assist with this endeavor, future research studies should provide more extensive information regarding the severity of participants' speech impairments, including reliable differential diagnoses of the participants' speech impairments.

Comprehensibility Versus Cognitive and Language Tests

Although it may seem obvious that the poor comprehensibility of children with severe speech impairments may negatively affect performance on standardized tests that

require spoken language (Beukelman & Mirenda, 2013; Glennen & DeCoste, 1997), even recent, relatively large-scale publications have reported language measures that rely on spoken communication for children with severe speech impairments (e.g., Vos et al., 2014). The same problem exists in clinical settings; all too often, clinicians such as pediatric neurologists, school-based diagnosticians and psychologists, and even speech-language pathologists administer tests that rely on spoken language to assess language and cognitive skills with children who have severe speech impairments. For example, the first author has received numerous reports from such professionals that mistakenly diagnosed children with language or cognitive impairments (some of whom participated in the current investigation), when, in reality, the children possessed only speech impairments. Such reports can lead to the provision of inappropriate services and school placements, which can affect long-term social and educational outcomes (Lund & Light, 2006).

In the current study, correlations between language sample comprehensibility versus several language and cognitive test scores were calculated to demonstrate that for children who have severe speech impairments with intact receptive language, significant correlations should not exist with test scores from tests that do not rely on spoken language. The three tests examined—TACL-3, PPVT-4, and LIPS-R—require no spoken language from the child during administration. As expected, comprehensibility scores were not significantly correlated with the participants' test scores. Future research that focuses on this issue should include a range of tests, including those that do and do not require spoken language, to fully demonstrate this crucial point. Careful consideration of test requirements is essential when administering standardized tests of any kind to children with severe speech impairments, and test interpretation must be undertaken with caution.

Summary

The findings provide initial evidence that language sampling may be used to extract important measures of expressive language for children with severe speech impairments. First, this study provides initial evidence that for some preschoolers with severe speech impairments, modifications can be made to traditional language sampling analysis to yield useful linguistic measures, such as MLUw and percentage comprehensibility, even when exact word agreement cannot be achieved. Once these methodologies are further validated, including children with severe speech impairments in studies of children with language impairments—who typically are excluded from such investigations (e.g., Rice et al., 2006, 2010)—may be possible. Second, comprehensibility from the language samples was correlated with single word comprehensibility from the I-ASCC, indicating that both approaches measured a similar construct for these participants. In a similar manner, MLUw from the transcripts was correlated with the mean of the three longest messages from the CDI (Fenson et al., 2006), again

indicating that these measures appear to be tapping into a similar construct. These findings lend support for using language samples to measure comprehensibility as well as collecting initial, broad syntactic data for some children with severe speech impairments. Last, comprehensibility from the language samples were not correlated with scores from the TACL-3, PPVT-4, or LIPS-R, indicating that the children with severe speech impairments may demonstrate a range of cognitive and linguistic abilities. This latter finding serves to caution against making assumptions about language or cognitive abilities on the basis of speech profiles (e.g., Rice et al., 1993). Further research is required to explore and validate the uses of language sampling for children with a range of severe speech impairments.

Acknowledgments

This research was supported by NIH Grant 1R03DC011610, awarded to the first author. A portion of this project served as a McNair Scholars research project for the second author. The authors thank Katherine Hustad for providing the initial inspiration for this project, Philip Dale and Amy Neel for assistance with early drafts of this article, Merissa Ekman for serving as our diligent second transcriber, and Lindsay Mansfield, Aimee Bustos, Esther Babej, Nathan Bickley, Bryan Ho, and Raleigh Kyabluu for additional assistance.

References

- Barefoot, S. M., Bochner, J. H., Johnson, B. A., Ann, B., & College, C. H. (1993). Rating deaf speakers' comprehensibility: An exploratory investigation. *American Journal of Speech-Language Pathology, 2*, 31–35.
- Barnes, E., Roberts, J., Long, S. H., Martin, G. E., Berni, M. C., Mandulak, K. C., & Sideris, J. (2009). Phonological accuracy and intelligibility in connected speech of boys with fragile X syndrome or Down syndrome. *Journal of Speech, Language, and Hearing Research, 52*, 1048–1061. doi:10.1044/1092-4388
- Beukelman, D., & Mirenda, P. (2013). *Augmentative and alternative communication: Supporting children and adults with complex communication needs* (4th ed.). Baltimore, MD: Brookes.
- Binger, C., Berens, J., Kent-Walsh, J., & Taylor, S. (2008). The effects of aided AAC interventions on AAC use, speech, and symbolic gestures. *Seminars in Speech and Language, 29*, 101–111. doi:10.1055/s-2008-1079124
- Binger, C., Kent-Walsh, J., Berens, J., Del Campo, S., & Rivera, D. (2008). Teaching Latino parents to support the multi-symbol message productions of their children who require AAC. *Augmentative and Alternative Communication, 24*, 323–338. doi:10.1080/07434610802130978
- Binger, C., Kent-Walsh, J., Ewing, C., & Taylor, S. (2010). Teaching educational assistants to facilitate the multi-symbol message productions of young students who require augmentative and alternative communication. *American Journal of Speech-Language Pathology, 19*, 108–120. doi:10.1044/1058-0360
- Binger, C., Kent-Walsh, J., & King, M. (in press). Dynamic assessment for three- and four-year old children who use augmentative and alternative communication: Evaluating expressive syntax. *Journal of Speech, Language, and Hearing Research*.
- Binger, C., Kent-Walsh, J., King, M., Webb, E., & Buenviaje, E. (in press). Early sentence productions of five-year-old children

- who use augmentative and alternative communication. *Communication Disorders Quarterly*. doi:10.1177/1525740116655804
- Binger, C., & Light, J.** (2007). The effect of aided AAC modeling on the expression of multi-symbol messages by preschoolers who use AAC. *Augmentative and Alternative Communication*, 23, 30–43. doi:10.1080/07434610600807470
- Carrow-Woolfolk, E.** (1999). *Test for Auditory Comprehension of Language* (3rd ed.). Austin, TX: Pro-Ed.
- Cook, B. G.** (2001). A comparison of teachers' attitudes toward their included students with mild and severe disabilities. *The Journal of Special Education*, 34, 203–213. doi:10.1177/002246690103400403
- Dowden, P.** (1997). Augmentative and alternative communication decision making for children with severely unintelligible speech. *Augmentative and Alternative Communication*, 13, 48–58.
- Dunn, L. M., & Dunn, D. M.** (2006). *Peabody Picture Vocabulary Test* (4th ed.). Bloomington, MN: AGS.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, S. J., & Bates, E.** (2006). *MacArthur–Bates Communicative Development Inventories User's Guide and Technical Manual* (2nd ed.). Baltimore, MD: Brookes.
- Flipsen, P.** (2006a). Measuring the intelligibility of conversational speech in children. *Clinical Linguistics & Phonetics*, 20, 303–312. doi:10.1080/02699200400024863
- Flipsen, P.** (2006b). Syllables per word in typical and delayed speech acquisition. *Clinical Linguistics & Phonetics*, 20, 293–301. doi:10.1080/02699200400024855
- Glennen, S. L., & DeCoste, D. C.** (1997). *Handbook of augmentative and alternative communication*. San Diego, CA: Singular.
- Gordon-Brannan, M., & Hodson, B. W.** (2000). Intelligibility/severity measurements of prekindergarten children's speech. *American Journal of Speech-Language Pathology*, 9, 141–150.
- Hodson, B. W., Scherz, J. A., & Stratman, K. H.** (2002). Evaluating communicative abilities of a highly unintelligible preschooler. *American Journal of Speech-Language Pathology*, 11, 236–242.
- Hustad, K. C., Allison, K., McFadd, E., & Riehle, K.** (2014). Speech and language development in 2-year-old children with cerebral palsy. *Developmental Neurorehabilitation*, 17, 167–175. doi:10.3109/17518423.2012.747009
- Hustad, K. C., Gorton, K., & Lee, J.** (2010). Classification of speech and language profiles in 4-year-old children with cerebral palsy: A prospective preliminary study. *Journal of Speech, Language, and Hearing Research*, 53, 1496–1513. doi:10.1044/1092-4388
- Hustad, K. C., Schueler, B., Schultz, L., & DuHadway, C.** (2012). Intelligibility of 4-year-old children with and without cerebral palsy. *Journal of Speech, Language, and Hearing Research*, 55, 1177–1190. doi:10.1044/1092-4388
- Kent-Walsh, J., Binger, C., & Buchanan, C.** (2015). Teaching children who use augmentative and alternative communication to ask inverted yes-no questions using aided modeling. *American Journal of Speech-Language Pathology*, 24, 222–236. doi:10.1044/2015_AJSLP-14-0066
- King, M., Binger, C., & Kent-Walsh, J.** (2015). Using dynamic assessment to evaluate the expressive syntax of five-year-old children who use augmentative and alternative communication. *Augmentative and Alternative Communication*, 31, 1–14. doi:10.3109/07434618.2014.995779
- Kwiatkowski, J., & Shriberg, L. D.** (1992). Intelligibility assessment in developmental phonological disorders: Accuracy of caregiver gloss. *Journal of Speech and Hearing Research*, 35, 1095–1104. doi:10.1044/jshr.3505.1095
- Lund, S. K., & Light, J.** (2006). Long-term outcomes for individuals who use augmentative and alternative communication: Part I—What is a “good” outcome? *Augmentative and Alternative Communication*, 22, 284–299. doi:10.1080/07434610600718693
- Lund, S. K., & Light, J.** (2007). Long-term outcomes for individuals who use augmentative and alternative communication: Part III—Contributing factors. *Augmentative and Alternative Communication*, 23, 323–335. doi:10.1080/02656730701189123
- McNeill, B. C., & Gillon, G. T.** (2013). Expressive morphosyntactic development in three children with childhood apraxia of speech. *Speech, Language and Hearing*, 16, 9–17.
- Miller, J. F.** (1981). *Assessing language production in children: Experimental procedures*. Needham Heights, MA: Allyn & Bacon.
- Miller, J. F., & Chapman, R. S.** (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, 24, 154–161. doi:10.1044/jshr.2402.154
- Monsen, R. B.** (1981). A usable test for the speech intelligibility of deaf talkers. *American Annals of the Deaf*, 12, 845–852. doi:10.1353/aad.2012.1333
- Mortimer, J., & Rvachew, S.** (2010). A longitudinal investigation of morpho-syntax in children with speech sound disorders. *Journal of Communication Disorders*, 43, 61–76.
- Murray, E., McCabe, P., & Ballard, K. J.** (2014). A systematic review of treatment outcomes for children with childhood apraxia of speech. *American Journal of Speech-Language Pathology*, 23, 486–504. doi:10.1044/2014
- National Center for Educational Statistics.** (2013). *Percentage distribution of students 6 to 21 years old served under Individuals with Disabilities Education Act (IDEA), Part B, by educational environment and type of disability: Selected years, fall 1989 through fall 2011*. Retrieved from https://nces.ed.gov/programs/digest/d13/tables/dt13_204.60.asp
- Owens, R. E.** (2012). *Language development: An introduction* (8th ed.). Boston: Pearson.
- Parker, M. D., & Brorson, K.** (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, 25, 365–376. doi:10.1177/0142723705059114
- Paul, R., & Norbury, C.** (2012). *Language disorders from infancy through adolescence: Listening, speaking, writing, and communicating* (4th ed.). St. Louis, MO: Elsevier.
- Rice, M. L., Hadley, P. A., & Alexander, A. L.** (1993). Social biases toward children with speech and language impairments: A correlative causal model of language limitations. *Applied Psycholinguistics*, 14, 445–471.
- Rice, M. L., Redmond, S. M., & Hoffman, L.** (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research*, 49, 793–808. doi:10.1044/1092-4388(2006/056)
- Rice, M. L., Smolik, F., Rytting, N., & Blossom, M.** (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research*, 53, 333–349. doi:10.1044/1092-4388(2009/08-0183)
- Roid, G. H., & Miller, L. J.** (1997). *Leiter International Performance Scale—Revised*. Wood Dale, IL: Stoelting.
- Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeney, J. L., . . . Wilson, D. L.** (2010). Extensions to the speech disorders classification system (SDCS). *Clinical Linguistics & Phonetics*, 24, 795–824. doi:10.3109/02699206.2010.503006
- Shriberg, L. D., Kwiatkowski, J., & Hoffmann, K.** (1984). A procedure for phonetic transcription by consensus. *Journal*

-
- of *Speech and Hearing Research*, 27, 456–465. doi:10.1044/jshr.2703.456
- Shriberg, L. D., & Lof, G. L.** (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics*, 5, 255–279.
- Shriberg, L. D., Paul, R., Black, L. M., & van Santen, J. P.** (2011). The hypothesis of apraxia of speech in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 41, 405–426. doi:10.1007/s10803-010-1117-5
- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A.** (2005). *Vineland Adaptive Behavior Scales* (2nd ed.). San Antonio, TX: Pearson.
- U.S. Census Bureau.** (2014). *State and county quick facts*. Retrieved from <http://quickfacts.census.gov/qfd/states/35000.html>
- van Dijk, M., & van Geert, P.** (2005). Disentangling behavior in early child development: Interpretability of early child language and its effect on utterance length measures. *Infant Behavior & Development*, 28, 99–117. doi:10.1016/j.infbeh.2004.12.003
- Viera, A. J., & Garrett, J. M.** (2005). Understanding inter-observer agreement: The kappa statistic. *Family Medicine*, 37, 360–363.
- Vos, R. C., Dallmeijer, A. J., Verhoef, M., Van Schie, P. E. M., Voorman, J. M., Wiegerink, D. J. H. G., . . . Becher, J. G.** (2014). Developmental trajectories of receptive and expressive communication in children and young adults with cerebral palsy. *Developmental Medicine & Child Neurology*, 56, 951–959. doi:10.1111/dmcn.12473
- Yorkston, K. M., & Beukelman, D. R.** (1981). Communication efficiency of dysarthric speakers as measured by sentence intelligibility and speaking rate. *Journal of Speech and Hearing Disorders*, 46, 296–301. doi:10.1044/jshd.4603.296
- Yorkston, K. M., Strand, E. A., & Kennedy, M. R. T.** (1996). Comprehensibility of dysarthric speech: Implications for assessment and treatment planning. *American Journal of Speech-Language Pathology*, 5, 55–66. doi:10.1044/1058-0360.0501.55