LSHSS

Research Article

Measuring Speech Production Development in Children With Cerebral Palsy Between 6 and 8 Years of Age: Relationships Among Measures

Phoebe Natzke,[a] Ashley Sakash,[a] Tristan Mahr,[a] and Katherine C. Hustad[a,b]

Purpose: Accurate measurement of speech intelligibility is essential for children with speech production deficits, but wide variability exists in the measures and protocols used. The current study sought to examine relationships among measures of speech intelligibility and the capacity of different measures to capture change over time.
Method: Forty-five children with cerebral palsy (CP) with and without speech motor impairment were observed at ages 6, 7, and 8 years. The speech performance of each child was rated using four measures at each time point: standardized articulation test scores, multiword intelligibility scores obtained from naïve listeners, parent ratings of intelligibility, and percent intelligible utterances obtained from language transcripts. We analyzed the correlations of measures within each age and within three different severity groups, and we analyzed how these measures changed year over year in each severity group.
Results: For children with CP who have mild and moderate speech deficits, different measures of speech production were weakly associated, and for children with CP with severe speech impairment, these measures showed stronger associations. The four measures also differed in their ability to capture change over time. Finally, results from standardized assessments of articulation were not found to inform overall speech intelligibility for children with mild and moderate speech deficits.
Conclusions: Results suggest that speech production is not fully described by any single clinical measure. In order to adequately describe functional speaking abilities and to capture change over time, multiple levels of measurement are required.

Children with cerebral palsy (CP) may present with a range of speech production problems stemming from different origins, including phonological and speech sound disorders and/or motor speech disorders. Previous work has shown that some children with CP demonstrate no detectible speech motor impairment (Hustad et al., 2010) while others, approximately 50% (Nordberg et al., 2013), have clinical evidence of dysarthria. Dysarthria ranges from mild to severe and manifests as deficits across any one or more speech subsystems, including articulation, resonance, phonation, and respiration (Darley et al., 1969). In addition, children with CP may demonstrate speech patterns or speech errors that are developmental in nature, either in isolation or as a comorbidity with dysarthria, further confounding the clinical assessment of speech performance and likely impacting decisions about targets for speech intervention (Allison & Hustad, 2018). Deficits in speech production, regardless of whether or not a child has dysarthria, often result in reduced speech intelligibility (Hustad et al., 2015; Namasivayam et al., 2013). The range of speech intelligibility deficits observed in children with CP is substantial (Hustad, Sakash, Natzke, et al., 2019). It is well documented that reduced speech intelligibility poses significant barriers to participation across environments (Dickinson et al., 2007; Ertmer & Goffman, 2011).

In academic contexts, children are expected to use spoken language to demonstrate knowledge, make requests, and engage with curriculum. Children also use spoken language to interact with peers, participate in recreational opportunities, and build social–emotional competencies. If a child is consistently not understood by peers and adults, that child may have a different and potentially diminished educational experience (Dickinson et al., 2007). Therefore,

[a]Department of Communication Sciences & Disorders, University of Wisconsin–Madison
[b]Waisman Center, University of Wisconsin–Madison
Correspondence to Phoebe Natzke: pnatzke@wisc.edu

Editor-in-Chief: Holly L. Storkel
Editor: Toby Macrae
Received November 4, 2019
Revision received January 28, 2020
Accepted April 6, 2020
https://doi.org/10.1044/2020_LSHSS-19-00102

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

a common treatment target for speech-language pathologists (SLPs) working with school-age children with speech impairments is the improvement of speech intelligibility (Hodge & Gotzke, 2014; Landa et al., 2014). In order to make decisions related to treatment and to track progress, some form of empirical measurement of speech intelligibility is essential (Kent et al., 1994; Williams et al., 2010). Speech intelligibility, however, is multidimensional and can be conceptualized and therefore assessed in a variety of ways (Connolly, 1986; Kent, 1993; Kent et al., 1994; Lagerberg et al., 2014).

In the school setting, standardized articulation measures are widely used to determine whether a child's acquisition or mastery of speech sounds is typical given their age. Articulation tests provide a simple and systematic approach to collecting a complete record of a child's ability to produce each speech sound across positions of word. Common to all tests of articulation is an inventory of speech sound production in single words. Some measures, such as the Goldman-Fristoe Test of the Articulation (Goldman & Fristoe, 2015), further assess speech sounds in connected speech. However, the degree to which scores from assessments designed to measure speech sound integrity can generalize to a description about overall speech intelligibility is unknown, and prior studies have yielded mixed results. Past research has indicated that segmental measures have weak relationships with transcription intelligibility (Ertmer, 2010). Although the number of articulation errors is generally negatively correlated with intelligibility, individuals can have significant articulation errors and still be highly intelligible (Whitehill, 2002). In this study, we examine how standardized articulation scores relate to other clinically practiced measures of speech intelligibility to determine whether these articulation scores can inform us about speech intelligibility.

In the motor speech domain, speech intelligibility refers to the ability of a speaker to successfully convey a message to a listener. This definition includes both the acoustic signal produced by the speaker and the ability of the listener to decode the signal (Kent et al., 1989; Yorkston & Beukelman, 1980; Yorkston et al., 1999, 1996). Measurement of speech intelligibility typically involves presenting speech samples of a known set of words and/or sentences produced by a speaker to one or more naïve listeners. Listeners, who are unfamiliar with the speaker and have limited context for the speech material, orthographically transcribe what they believe the speaker said. A percent intelligible score is obtained by computing the number of words correctly transcribed, divided by the total number of words produced. In this paradigm, production accuracy of individual speech sounds is not considered; instead, the accuracy of individual words making up the sample is the unit of measure. This approach is highly objective and is considered by some to be a gold standard (Lagerberg et al., 2014). For frontline SLPs, however, this approach is time consuming and requires resources not often available in clinical, school, or home environments. In addition, because there is no communicative context, results may underestimate speech intelligibility in real-life communication situations.

Another method of characterizing speech intelligibility uses data from the analysis of language samples. Although the primary purpose of language transcription and associated language sample analysis is to assess oral language skills (Heilmann et al., 2010), clinicians and researchers have used language sample analysis data to describe intelligibility of children's utterances (Binger et al., 2016; Rice et al., 2010; Yoder et al., 2016). Specifically, the measure percent intelligible utterances (PIU) indicates how many of the utterances within a specific sample were fully intelligible to the transcriber. In this context, the unit of measure is the whole utterance, and the integrity of individual words and the phonemes that make up each utterance is not taken into account. For instance, an utterance containing five words with one word that is unintelligible is scored the same as an utterance in which all five words are unintelligible; the utterance is considered wholly unintelligible in both cases. In language sample analysis, a highly trained individual who has expertise in speech-language pathology completes the transcription. As in most natural interactions, the transcriber is aided by contextual clues provided by the elicitation procedure, the communication partner's responses to child utterances, glosses by the partner, and the environment in which the sample was collected. As a result, there is considerable context that supports the transcriber's ability to make sense of the child's productions. Thus, this direct measurement approach is more holistic and less granular in nature than orthographic transcription of decontextualized utterances by unfamiliar listeners. In research contexts, transcribers are often allocated a limited number of attempts to transcribe an utterance, marking the vocalization as unintelligible if the full utterance cannot be transcribed after three repetitions (Binger et al., 2016; Hustad et al., 2014). Recent work has demonstrated that it is possible to collect and reliably transcribe language samples from children with speech impairments (Binger et al., 2016; Yoder et al., 2016), and transcription reliability has been reported as ranging from acceptable to very good, even for children for whom half of their utterances are unintelligible (Yoder et al., 2016). Even though a valuable feature of language sample analysis is the ecological validity of measures obtained from natural conversational contexts, one drawback of this method is that the speaker's intended target is not known. That is, the content of the language sample is accepted as accurate if the transcriber assigned words to the child's spoken message—a more accurate name for the resulting measure might be percent transcribable utterances.

An indirect method for obtaining information about a child's ability to effectively use speech and to be understood involves asking parents and SLPs to make perceptual judgments of speech intelligibility. Parent and educator report is a commonly used method in both school and research settings. Rather than trying to identify specific items (phonemes, words, or utterances), raters score on a speaker's intelligibility either on a numeric scale (e.g., a 5-point scale where 1 indicates *speech that is not at all understood* and 5 indicates *almost perfect speech*) or by selecting descriptive categories (e.g., "mildly impaired" vs. "severely impaired"). There are a variety of published rating scales, including the Viking Speech Scale (Pennington et al., 2013) and the Intelligibility in Context Scale (ICS; McLeod et al., 2012). Researchers and clinicians also make use of informal rating scales, which are presented to parents or professionals as simple interview questions or categorical ratings (Bickham et al., 2017; Fox & Boliek, 2012; Malmenholt et al., 2019). These types of ratings, whether informal or from a published scale, are very holistic in nature because they ask raters to consider speech intelligibility broadly, outside any specific speech sample. Perceptual ratings are indirect measures and can provide significant time savings compared to direct measures where specific speech samples are objectively analyzed in some way. Perceptual rating scales, such as the ICS, also permit examination of how a child's speech intelligibility differs across environments, allowing determination of where and when a child's participation is hindered by reduced intelligibility (McLeod et al., 2012). However, drawbacks to using parent and professional ratings to characterize speech intelligibility include the subjective nature of ratings and the possibility that coarsely scaled ratings may not capture small changes in performance (Ertmer, 2010).

Clinical judgment is required when making decisions about how to accurately and comprehensively characterize the speech production abilities of an individual child. Clearly, the methods described above all yield useful information from different perspectives (i.e., speech sound, word, utterance, holistic perception), but the extent to which each measure is related to the others, the effects of severity on these relationships, and whether the measures capture change over time is unknown. Given both the evidence that neurological changes can and do compromise the integrity of all speech subsystems and the wide range of speech production abilities observed in children with CP, finding answers to these questions is especially critical for SLPs who serve children in this population. For example, a child with intact articulatory abilities may struggle with generating enough breath support for intelligible speech. A standardized test of articulation would result in a score within normal limits for this child, but this finding would be contradictory to parent or teacher report that the child struggles to communicate in noisy environments. So while a measure of articulation is not an unreasonable selection for a child who would be identified as having a high risk of speech sound errors (particularly distortions), this child might inaccurately be identified as not needing the services of an SLP. This example illustrates that any of the tools described above are reasonable selections when seeking to understand more about a child's functional communication abilities, but having a deeper understanding of commonalities and/or differences between evaluative materials is critical for ensuring that our clinical assessments are truly comprehensive and that treatments are designed to address the full range of communication needs observed in children with CP.

In this study, we examined the interrelations among different measures of speech intelligibility and speech production to begin to understand how well inferences obtained from one type of measure might be related to outcomes on other speech production measures. Furthermore, we examined how these relations differed based on severity of the speech impairment and how the measures themselves changed with growth over time. Specific questions were as follows:

1. What are the relationships among different measures of speech production for children with CP within each of three severity groups? Are relationships similar among measures between severity groups?
2. How do each of the measures change over time within severity groups? Are the different measures sensitive to change within severity groups and are there differences between groups in the magnitude of change for the different measures?

We hypothesized that there would be differences in the relationships observed between measures based on severity. We expected that relationships between measures would correlate more strongly in children with more severe speech impairment because the presence of more pervasive deficits was expected to impact speech production broadly, regardless of how granular the level of measurement is. Similarly, we expected that the associations would be weaker and less predictable for children with mild or moderate speech impairment, because context cues and coarser measurement units can compensate for these children's speech errors. Finally, with regard to change over time, we expected each of the measures to show improvement with age within each severity group.

Method

Participants

Children With CP

Forty-five children with CP (23 boys, 22 girls) participated in this project. Participants were selected from a larger cohort of children recruited for a longitudinal study on the communication development of children with CP, which was approved by the University of Wisconsin–Madison Institutional Review Board for Social and Behavioral Sciences. Eligibility for the larger research study included the following: (a) medical diagnosis of CP and (b) hearing within normal limits as documented by formal audiological evaluation or distortion product otoacoustic emission screening. For this study, each participant was required to meet the following additional inclusion criteria: (c) completed three data collection sessions, one each at ages 6, 7, and 8 years; (d) ability to produce, at a minimum, one- and two-word stimuli from the Test of Children's Speech (TOCS+; Hodge & Daniels, 2007). In total, ninety-nine children with CP made visits to the study between the ages of 6;0 and 8;11 (years;months). Forty-four of those children were excluded from the current study because they were anarthric and did not complete the one- and two-word stimuli from the TOCS+. Ten additional children had produced the required TOCS+ stimuli but were excluded as they did not complete a visit at each of the three required time points. The resulting 45 children were included in the analysis for this current study. The average age of the children at the 6-year-old visit was 77.6 months (*SD* = 3.8), the average age of the children at the 7-year-old visit was 89.1 months (*SD* = 4.5), and the average age of the children at the 8-year-old visit was 100.9 months (*SD* = 4.6). The 45 children included in this study contributed 135 total visits. Table 1 provides demographic information about the child participants.

Children in this study demonstrated a range of speech production abilities. Thirty-one of 45 children had clinical evidence of speech motor impairment, specifically, dysarthria. The presence of dysarthria was determined by the first two authors (both SLPs) with expertise in pediatric dysarthria, using clinical judgment. Each SLP assessed audio recordings of the speech of each child, including productions of single-word and multiword items from the TOCS+ (Hodge & Daniels, 2007) and video recordings of a parent–child interaction and a clinician–child interaction. Using the recordings, children were evaluated for perceptual features of speech suggestive of articulatory (e.g., reduced speech rate, imprecise consonants, distortions, omission, substitutions that were not age appropriate), resonatory (e.g., hypernasality, nasal air emission), phonatory, and/or respiratory (short breath groups, hoarse or harsh voice, breathy voice, low vocal volume) involvement, which would indicate the presence of speech subsystem impairment. These same sources were also examined for evidence of drooling, facial asymmetry at rest and during movement, and the presence of abnormal tone in the orofacial musculature. Following examination of both these sources of information, the SLPs triangulated their observations and made a binary determination on the presence or absence of dysarthria for each child at the time of their 6-year-old visit. Agreement on the presence/absence of dysarthria was 100%.

Table 1. Demographic characteristics of participants with cerebral palsy by severity group.

Characteristic	Mild *n* = 17	Moderate *n* = 15	Severe *n* = 13
Male:female ratio	11:6	7:8	5:8
M_{age}, months (*SD*)			
Time 1	79 (4.1)	76. (3.1)	78 (3.7)
Time 2	91 (4.3)	87 (3.6)	89 (5.1)
Time 3	102. (4.9)	100 (3.9)	101 (4.8)
Mean Arizona-3 SS Time 1	89.29 (7.86)	85.71 (9.1)	69.69 (10.4)
Mean Arizona-3 SS Time 2	88.64 (5.37)	88.27 (7.35)	70.08 (11.34)
Mean Arizona-3 SS Time 3	90.53 (5.4)	88.54 (8.78)	72.25 (12.3)
Type of CP			
Spastic	14	13	8
Diplegia	4	4	
Hemiplegia (left)	7	2	1
Hemiplegia (right)	3	4	2
Triplegia			1
Quadriplegia			4
Dyskinetic		1	
Ataxic	2	1	2
Mixed			1
Hypotonic	1		
Unknown			2
GMFCS at age of 6 years			
I	5	5	6
II	7	8	6
III	3		1
IV		2	
V	2		

Note. The Gross Motor Function Classification System (GMFCS) is a tool that allows for the categorization of the gross motor function of children and young people with cerebral palsy (CP); providing clinicians with a description of motor function and a sense of what equipment and mobility aids are required (Palisano et al., 2008). Arizona-3 SS = Arizona Articulation Proficiency Scale–Third Edition standard score.

For the 14 children who did not have clinical evidence of speech motor impairment, two had standard scores on the Arizona Articulation Proficiency Scale–Third Edition (Arizona-3; Fudala, 2001) that were outside the range of typical. The rest had typically developing articulation.

Children were categorized into three severity groups on the basis of speech intelligibility scores obtained from single-word productions at the 6-year-old visit. There is a long history in the dysarthria literature of assigning severity designations on the basis of intelligibility scores (Weismer & Martin, 1992). Although there is not a single set of intelligibility guidelines for assigning severity designations, we selected criteria based on our earlier work with children with CP (Hustad, Sakash, Natzke, et al., 2019). In addition, use of single-word intelligibility scores to create severity groups preserved the integrity of our focus on intelligibility while also ensuring that all of our key dependent variables were preserved. We chose not to group children by dysarthria status, because our previous work has shown that intelligibility reductions are not strictly dependent on dysarthria status in children who have CP. That is, children with CP without dysarthria may present with reduced intelligibility, and children with CP who have dysarthria may present with very high intelligibility (Hustad, Sakash, Broman, & Rathouz, 2019). Second, children in the sample varied in the lengths of utterance they were able to produce, but single-word productions were obtained from all participants. Using single-word intelligibility scores, we operationally defined groups as follows: children with single-word intelligibility scores up to 60% made up the severe group; children with single-word intelligibility scores between 61% and 80% made up the moderate group; children with single-word intelligibility scores at or above 81% made up the mild group. In the severe group, there were 13 children; all had clinical dysarthria. In the moderate group, there were 15 children; 12 had clinical dysarthria. In the mild group, there were 17 children; six had clinical dysarthria. Although it might be expected that all children without clinical dysarthria would score into the mild severity group and that only children with dysarthria score into the moderate and severe groups, the variability in groupings is consistent with our previous studies.

Naïve Adult Listeners

A total of 270 naïve adult listeners (two different listeners per child and per visit: 45 children × 3 visits × 2 listeners = 270) provided orthographic transcriptions of children's speech, which were used to quantify intelligibility for several of the measures used in this study. Listeners were recruited via public postings and online advertisements; they were compensated for their participation. Listeners were between 18 and 45 years of age, identified English as their first language, and reported a negative history for brain injury or speech/language/cognitive disability. Each listener was required to pass a pure-tone hearing screening at 25 dB HL for 250 Hz, 500 Hz, 1 kHz, 4 kHz, and 6 kHz bilaterally. Listeners were unfamiliar with the children who participated in the study. Of the participants, 221 women and 49 men participated as listeners, and the mean age was 21.24 years (SD = 4.08).

Materials and Procedure

During each diagnostic session, each child completed a battery of speech, language, and oral–motor assessments (see Hustad et al., 2010). In addition, parents completed informal questionnaires about various aspects of their child's development. All sessions were administered by a certified SLP in a sound-attenuating suite. Individual sessions were up to 3 hr in duration and were tolerated without difficulty. All speech produced by the children during each diagnostic session was recorded using a digital audio recorder (Marantz PMD 570) at a 44.1-kHz sampling rate (16-bit quantization). A condenser studio microphone (Audio-Technica AT4040) was positioned next to each child using a floor stand and was located approximately 18 in. from the child's mouth. Throughout the session, the level of the signal was monitored and adjusted on a mixer (Mackie 1202 VLZ) to obtain optimized recordings and to avoid peak clipping.

In this study, we examined four different measures: multiword intelligibility scores obtained from naïve listeners, parent ratings of how understandable they thought their children were to others, standardized articulation test scores, and PIU obtained from language transcripts. We obtained these measures via structured and unstructured tasks. In the sections that follow, we describe the method used in obtaining and analyzing each of these four measures.

Multiword Transcription Intelligibility

Each child completed a single-word and sentence imitation task at each visit. This task included stimuli from the TOCS+ (Hodge & Daniels, 2007). For these measures, an iPad was used to present each child with an image and a prerecorded auditory model, which was immediately repeated by the child. The stimulus set for this study included 42 individual words and 60 sentences ranging from two to seven words (10 of each sentence length). Lexical, phonetic, syntactic, and morphological features of all stimuli were developed to be appropriate for children. Note that not all children were able to produce utterances of each sentence length due to speech motor constraints (see Table 2).

Digital recordings of each child's production of stimuli from the TOCS+ were separated into single audio files and were peak amplitude normalized to ensure that maximum loudness levels were the same across all children and all utterances. This procedure also ensured that the amplitude contours of the original productions remained preserved and enabled calibration to predetermined output levels for playback to listeners. Audio files were presented to adult listeners using in-house software in a sound-attenuating booth. Listeners completed two orthographic transcription tasks, one involving single words and one involving multiword utterances. The order of presentation of the two tasks was counterbalanced across the two listeners for each child; individual utterances within each task were randomized for each listener. Listeners were instructed that all productions were of real words; they heard each utterance one time prior to transcribing what they heard. Intelligibility was determined by comparing listeners' orthographic transcriptions against the target utterances that children produced. Each word typed by listeners was scored as either correct or incorrect based on whether it matched the target word produced by the child phonemically. Homonyms and misspellings were accepted as correct if the phonemes from the spoken version of the utterance matched what the listener typed. Each child's visit had two different listeners.

For each child's visit, we computed the difference in the percentage of words transcribed correctly between the two listeners (Lee et al., 2014). For example, if the percentage of words transcribed correctly was 90% for Listener 1 and 82% for Listener 2, this would be an 8% difference. The resulting overall intelligibility score would be 86%.

Table 2. Number of children who produced each sentence length by severity group and age.

Sentence length	Age (years)	Mild (n = 17)	Moderate (n = 14)	Severe (n = 14)
2 words	6	17	14	14
	7	17	14	14
	8	17	14	14
3 words	6	17	14	13
	7	17	14	14
	8	17	14	14
4 words	6	17	14	11
	7	17	14	11
	8	17	14	14
5 words	6	16	14	9
	7	17	14	9
	8	17	14	10
6 words	6	16	14	6
	7	17	14	8
	8	17	14	9
7 words	6	16	14	6
	7	16	14	8
	8	17	14	9

However, if the two listeners' scores differed by more than 10 percentage points (e.g., 60% for Listener 1 and 45% for Listener 2), data from a third listener was obtained and data from the listener who differed from the other two by more than 10 percentage points was discarded. Of the 135 child visits, this occurred in 12 instances. For the final data set used for this study, the average difference between the two listeners was 3.5 percentage points (3.1 *SD*s). These values are well within the range of variability deemed acceptable following Lee et al. (2014).

We also calculated interrater reliability of the average multiword intelligibility scores for the two listeners used for each child and each visit using the intraclass correlation coefficient (ICC) estimated with the irr R package (Version 0.84.1; Gamer et al., 2019). We used an average score, consistency-based, one-way random effects model, and we observed strong agreement between raters on intelligibility scores, ICC = .994, 95% CI [.991, .995].

Intelligibility scores were obtained as follows: The total number of words identified correctly across the two listeners for all utterances produced by a given child was divided by the total number of words possible (across the two listeners) and multiplied by 100 to yield a percent intelligibility score for each child at each time point. Speech intelligibility scores for single words were used to create the severity groups described above. Speech intelligibility scores for multiword utterances were used as a dependent variable in this study.

Parent Ratings of Understandability

At each visit, the parent accompanying the child completed an informal communication questionnaire designed to gather information regarding parent perceptions of their child's receptive and expressive language abilities as well as their child's ability to effectively use speech across environments. Questions contained in this questionnaire addressed the child's functional speaking abilities with both familiar and unfamiliar communication partners. In the current study, we examined the responses to the question: "Overall, how understandable is your child to others (even if he/she doesn't sound 'normal')?" Parents were instructed to encircle a number from 1 to 7, where a rating of 1 indicated *very easy to understand* and a rating of 7 indicated *very hard to understand*. Parent response data were missing for four of 135 visits.

Standardized Articulation Assessment

Each child participated in a standardized articulation assessment at each visit. We used the Arizona-3 because it assesses each sound in Standard American English, including vowels, in all allowable positions of words. Following the test manual, children were presented with picture cards of familiar objects and asked to name each item. Scoring for each sample was completed by a research assistant (a graduate student in speech-language pathology), using audio recordings of the child's responses. The research assistant made binary (correct/incorrect) judgments for each phonetic item following instructions from the test manual, and each child received a raw score.

Because of the known difficulties related to transcription of dysarthric speech, we quantified interrater reliability on all Arizona-3 data. That is, we had a second research assistant score each Arizona-3 administration for each child, and we computed an ICC score using a single-score, absolute-agreement, two-way random effects model (Gamer et al., 2019). We found strong agreement between the two raters, ICC = .907, 95% CI [.682, .959]. We analyzed raw scores because of the narrow age range of the children in the sample for each visit under the assumption that raw scores would provide finer grained information on growth of speech sound accuracy over time than standard scores. Note that five of 135 visits did not include administration of the Arizona-3, and therefore, data were missing.

Intelligible Utterances From Language Samples

Each child participated in a parent–child play-based interaction during each visit. A standard set of play items were provided for all participants, and parents were instructed to play and talk to their children as they normally would. Using standard Systematic Analysis of Language Transcripts (SALT) conventions, parent–child interactions were transcribed. SALT conventions allow for systematic segmentation of utterances and notation of segments of speech that are deemed unintelligible. For this study, utterances were segmented in communication units (Loban, 1976; Miller et al., 2011), and to determine and demarcate unintelligible segments of speech, transcribers were instructed to listen to the child's verbal production up to 3 times. Following the third pass, words or utterances that were not understood by the transcriber were marked as "XXX." Due to the presence of dysarthria, no attempt was made to segment or quantify the number of words per unintelligible segment. Traditional guidelines indicate that clinicians should transcribe a fixed quantity of material, commonly 50 complete and intelligible child utterances, as a representative sample of a child's expressive language ability (Heilmann et al., 2010). However, as indicated by previous research (Hustad et al., 2014), both the total number of utterances and the efficiency with which these utterances were produced (rate of communication) were anticipated to be highly variable for this cohort of children with CP, and accordingly, we standardized the duration of parent–child interactions to 9 min 40 s (the duration of the shortest interaction). We included all utterances that occurred within this time frame whether or not they exceeded 50 utterances.

From the transcripts of the parent–child play interactions, we computed the measures of "total utterances" (defined as the number of utterances in the transcript, including those that contained an unintelligible segment of speech) and "intelligible utterances" (defined as the number of utterances that contained no unintelligible segments of speech). Note that utterances that were abandoned or interrupted were included in the analysis set as the primary goal was to examine intelligibility as measured via language transcription. The PIU—the primary measure of interest obtained from the transcripts—was defined as the number of intelligible utterances divided by the total number of utterances (and multiplied by 100), based on the entire transcript.

To ensure that transcription-based data were reliable, interaction samples were randomly selected from 16 different children and were independently transcribed by a second trained transcriber. We then compared the proportions of intelligible utterances between raters. We computed an ICC score using a single-score, absolute-agreement, two-way random effects model (Gamer et al., 2019), and we observed good agreement between the two raters, ICC = .802, 95% CI [.530, .925]. Note that at two of 135 visits, the child participant did not engage in a parent–child play interaction, and therefore, data were missing.

Experimental design and statistical procedures. Research questions of interest focused on examining (a) relationships among the following measures of speech production (Arizona-3 scores, multiword intelligibility scores, PIU from language transcripts, parent ratings of intelligibility) within severity groups and within different age points (6, 7, and 8 years) and (b) statistical differences over the three age points within severity groups for each of the four measures. To examine relationships between variables within severity groups and time points, we used nonparametric Kendall's tau correlation coefficients. We partitioned alpha using the Bonferroni procedure. For each family of correlations, a p value of .0028 or less was required for significance (alpha = .05 allocated to each family; .05/18 correlations = .0028).

A 1 × 3 design was employed to examine differences over age points (6, 7, and 8 years) within each of the three severity groups (mild, moderate, severe) on each of the four dependent measures. Data were longitudinal in nature for each child; thus, one-way repeated-measures analyses of variance were performed, along with dependent-samples follow-up contrasts to examine change over time on each measure. We used the nonparametric Friedman's analysis of variance because of the small within-group sample size and because several of our measures were ordinal in nature. Again, we used the Bonferroni procedure to partition alpha. For each omnibus test, a p value of .0167 or less was required for significance (alpha = .05 allocated to each family; .05/3 severity levels per family = .0167). To examine pairwise differences among age levels for significant omnibus results, we used Wilcoxon signed-rank tests. For each follow-up contrasts, a p value of .008 or less was required for significance (alpha = .05 allocated to each family; .05/6 contrasts = .008).

Results

Relationships Among Measures Within Severity Groups and Age Levels

Table 3 shows descriptive results for each of the measures by age and severity group. Table 4 provides nonparametric correlations between measures by time and severity group. Results showed that, for children in the mild group, none of the correlations among measures was significant at any of the age points 6, 7, and 8 years. Similarly, none of the correlations among measures was significant for children in the moderate group for any of the age points.

Table 3. Descriptive results for Arizona Articulation Proficiency Scale–Third Edition (Arizona-3) raw scores, percent intelligible utterances (PIU) from language transcripts, multiword intelligibility scores (MWI) based on unfamiliar listener transcriptions, and parent ratings of intelligibility to others (PR-others) by severity group and age.

Measure	Age (years)	Mild *M*	Mild *SD*	Mild *n*	Moderate *M*	Moderate *SD*	Moderate *n*	Severe *M*	Severe *SD*	Severe *n*
Arizona-3	6	94.32	4.63	17	89.89	7.69	14	69.04	18.85	13
	7	95.94	4.19	17	95.07	5.06	15	74.08	19.15	12
	8	97.59	4.05	17	96.42	6.06	13	80.54	16.16	12
PIU	6	86.79	6.06	17	86.50	7.65	15	66.03	20.69	13
	7	89.63	4.92	17	88.62	7.72	15	73.30	18.25	11
	8	94.74	4.75	17	94.29	6.74	15	71.73	21.23	13
MWI	6	87.40	10.08	17	75.28	10.52	15	25.30	20.06	13
	7	92.14	5.54	17	80.79	11.19	15	27.55	18.66	13
	8	94.76	3.89	17	83.77	12.10	15	37.22	23.06	13
PR-others	6	1.81	1.05	16	3.00	1.31	15	4.62	1.50	13
	7	1.88	1.02	16	3.50	1.65	14	4.62	1.56	13
	8	2.00	1.41	16	3.13	1.51	15	4.46	1.33	13

Table 4. Kendall's tau nonparametric correlation coefficients by age and severity group.

Age (years)	Severity	Measure	Arizona-3	PIU	MWI
6	Mild	PIU	−.04		
		MWI	.14	−.03	
		PR-others	−.36	−.17	−.497
	Moderate	PIU	.18		
		MWI	.11	−.07	
		PR-others	−.441	−.568	−.05
	Severe	PIU	.564		
		MWI	.710*	.529	
		PR-others	−.42	−.36	−.41
7	Mild	PIU	.16		
		MWI	.20	.01	
		PR-others	−.13	−.21	−.433
	Moderate	PIU	.12		
		MWI	.37	−.04	
		PR-others	−.513	−.08	−.451
	Severe	PIU	.42		
		MWI	.718*	.491	
		PR-others	−.502	−.19	−.595
8	Mild	PIU	.409		
		MWI	.04	−.04	
		PR-others	−.12	−.19	−.20
	Moderate	PIU	.06		
		MWI	.13	−.11	
		PR-others	−.473	.10	−.562
	Severe	PIU	.687*		
		MWI	.748*	.513	
		PR-others	.10	−.17	−.06

Note. Arizona-3 = Arizona Articulation Proficiency Scale–Third Edition; PIU = percent intelligible utterances; MWI = multiword intelligibility scores; PR-others = parent rating of intelligibility to others.

*$p < .0028$.

Finally, for children in the severe group, the correlation between multiword intelligibility and Arizona-3 score was significant and strong at each of the age points (taus = .710, .718, .748; $p < .0028$, for the three age points, respectively). In addition, the correlation between PIU and Arizona-3 scores was significant with a moderate–strong effect at 8 years of age for children in the severe group (tau = .687, $p < .0028$).

Differences Over Time Within Severity Group on Each Measure

Figures 1–4 show longitudinal change over time for each of the three severity groups of children on each of the four measures. Descriptive results for group data suggest that there was steady improvement on each of the measures over time, except for parent ratings of intelligibility, which showed inconsistent findings. Inferential statistics, shown in Tables 5 and 6, revealed that, for children in the mild group, there was significant change over time in multiword intelligibility scores ($\chi^2 = 14.24$, $p < .001$) and in PIU ($\chi^2 = 16.35$, $p < .001$). Follow-up contrasts indicated that the difference between 6- and 8-year time points was significant for multiword intelligibility scores ($Z = -1.29$, $p < .001$) and for PIU ($Z = -1.35$, $p < .001$). In addition, the differences between 7- and 8-year time points was significant for PIU ($Z = -.94$, $p = .006$).

For children in the moderate group, there were significant changes over time in Arizona-3 scores ($\chi^2 = 13.14$, $p = .001$) and multiword intelligibility scores ($\chi^2 = 9.73$, $p = .008$). Follow-up contrasts revealed that the difference between 6- and 8-year time points was significant for both Arizona-3 scores ($Z = -1.417$, $p = .001$) and for multiword intelligibility scores ($Z = -1.133$, $p = .002$).

For children in the severe group, there were significant changes over time only in Arizona-3 scores ($\chi^2 = 16.55$, $p < .001$). Follow-up contrasts revealed that the difference between 6- and 8-year time points was significant for Arizona-3 scores ($Z = -1.417$, $p = .001$).

Discussion

This study examined the interrelations between four measures of speech production as well as change over time on each measure for three groups of children with CP that varied in the severity of their speech involvement. Variables of interest differed in their level of measurement and included indices of speech sound integrity (phoneme level), intelligibility of connected speech as measured by unfamiliar listeners (lexical level), PIU as obtained from language transcripts (utterance level), and parent ratings of how intelligible they believe their child is to others (holistic level). There were two broad sets of findings from this study. First, relationships among measures differed by severity group, with the only significant correlations observed within the group of children with the most severely reduced intelligibility. Second, not all variables showed consistent growth over time for all severity groups, and when growth was observed, it tended to be evident for 6 versus 8 years of age rather than at yearly intervals. These findings and their implications for assessment of children's speech are discussed below.

Figure 1. Arizona Articulation Proficiency Scale–Third Edition (Arizona-3) raw scores, by severity group, from 6 to 8 years of age. *p < .0083.

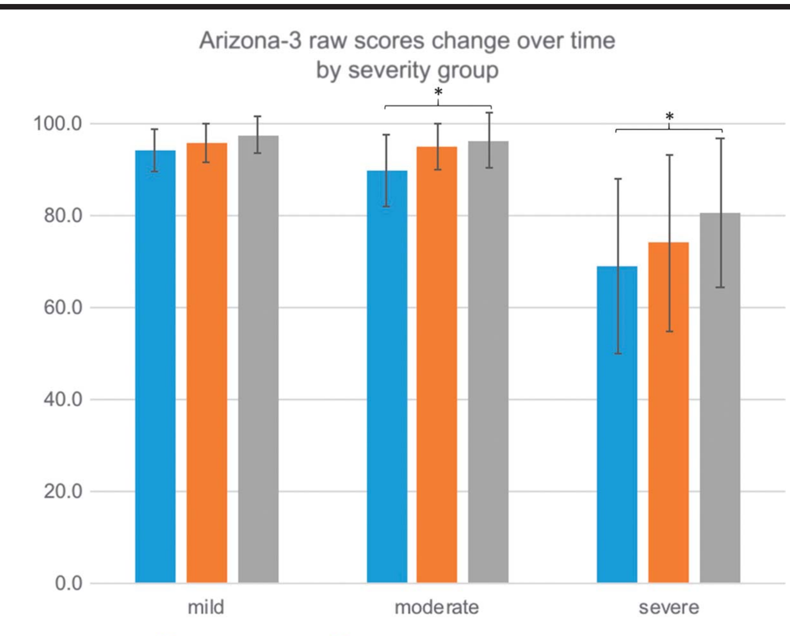
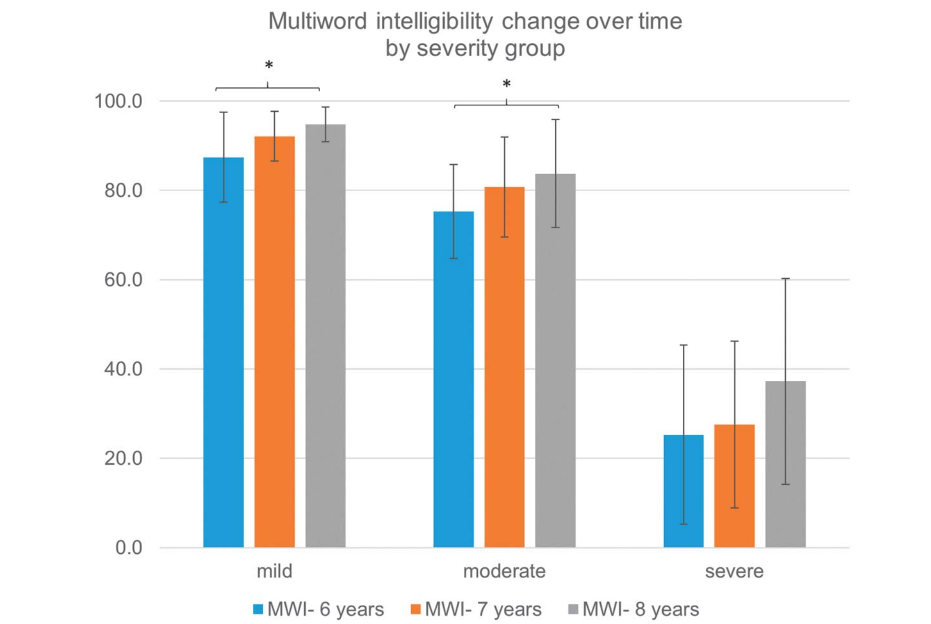


Figure 2. Multiword intelligibility scores, by severity group, from 6 to 8 years of age. *p < .0083.



Relationships Among Measures Within Severity Groups and Age Levels

Results of this study showed that within the mild and moderate severity groups, none of the measures was significantly correlated, suggesting that each measure was capturing something unique and unrelated to the other measures. Specifically, segmental integrity (Arizona-3 raw scores) was unrelated to any measure of speech intelligibility, and none of the measures of speech intelligibility was related to each other. Notably, these findings suggest that measures at the level of the speech sound, the individual word, the utterance, and the overall perception of intelligibility are not reflective of one another in a consistent

Figure 3. Percent intelligible utterances (PIU) by severity group, from 6 to 8 years of age. *p < .0083.

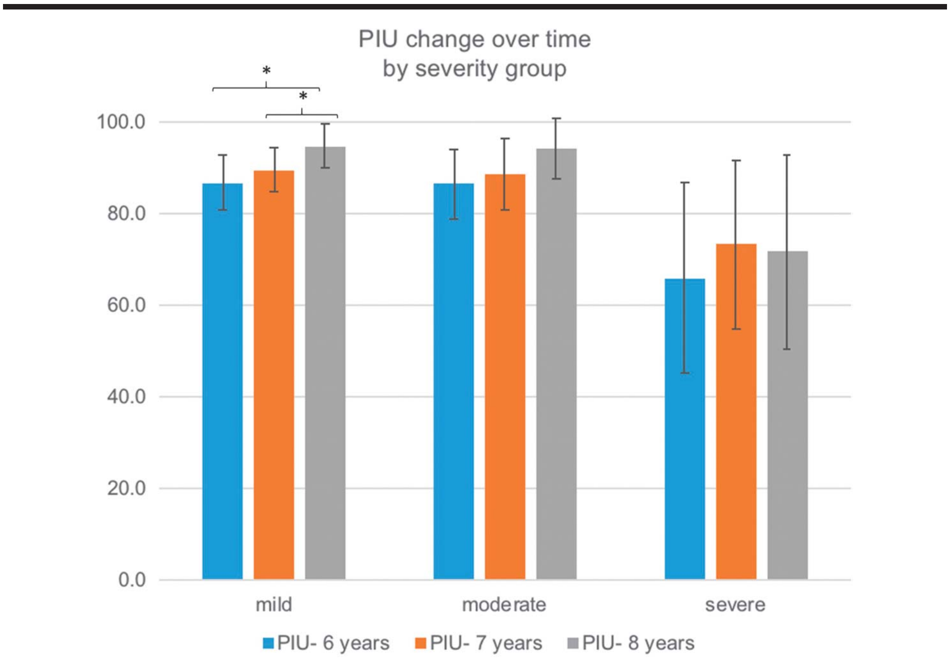
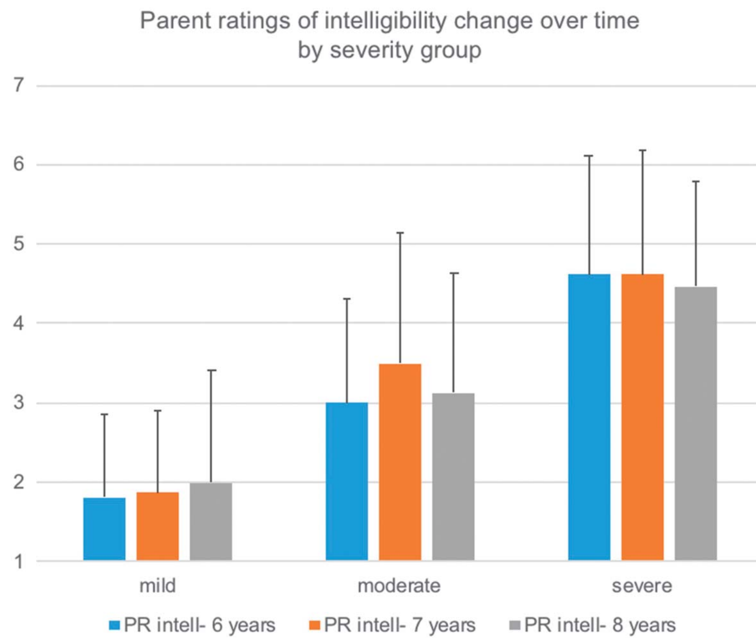


Figure 4. Parent ratings of intelligibility (PR intell), by severity group, from 6 to 8 years of age. Note that parents were asked to describe their perception of the intelligibility of their child's speech using a scale from 1 to 7, where 1 is the best rating (*very easy to understand*) and 7 is the worst rating (*very hard to understand*).



and meaningful way for children with single-word intelligibility above 61%. From a clinical perspective, these results suggest that none of the measures examined in this study can serve as a proxy for any of the other measures for children with mildly or moderately reduced intelligibility. That is, we cannot generalize from articulation to any level of intelligibility, nor can we generalize from intelligibility to articulation. Intelligibility (regardless of level—word, utterance, overall) and speech sound integrity must be measured explicitly. This finding has important implications for interpretation of literature in domains such as child language, where PIU is often used as an index of intelligibility or speech sound development/disorders, where inferences are often made about intelligibility based on segmental results. Results of this study suggest that findings from PIU and from segmental development, while valid and legitimate in and of themselves, do not yield the same information as other measures of intelligibility, most notably lexical-level intelligibility that is commonly used in the speech motor literature. We consider this finding further in the context of change over time in each measure. One key message is that, to fully understand speech performance at all levels, measurement should be conducted at all levels.

Table 5. Related-samples Friedman's analysis of variance by ranks omnibus test results within severity group over time for each measure.

Source	*n*	*df*	χ^2	*p*
Arizona-3				
Mild	17	2	6.64	.036
Moderate	15	2	13.14	.001*
Severe	11	2	16.55	< .001*
MWI				
Mild	17	2	14.24	.001*
Moderate	15	2	9.73	.008*
Severe	13	2	3.85	.146
PIU				
Mild	17	2	16.35	< .001*
Moderate	15	2	6.93	.031
Severe	11	2	2.36	.307
PR-others				
Mild	16	2	.621	.733
Moderate	14	2	1.08	.581
Severe	13	2	1.12	.572

Note. Arizona-3 = Arizona Articulation Proficiency Scale–Third Edition; MWI = multiword intelligibility scores; PIU = percent intelligible utterances, PR-others = parent rating of intelligibility to others.

*$p < .0167$.

One important finding from this study is that interrelations among measures were different for children in the severe group. For these children, segmental integrity (Arizona-3 scores) was correlated with multiword intelligibility scores (lexical level) at all three time points. Arizona-3 scores were also correlated with PIU (sentence level) at 8 years of age. This finding suggests that reduced lexical intelligibility is closely related to reduced segmental integrity for children with severe intelligibility deficits, an observation that is not surprising. Specifically, if a child has a severe speech impairment that has a marked impact on production

Table 6. Wilcoxon signed-rank test follow-up contrasts for significant omnibus tests.

Contrast	*SE*	*Z*	*p*
Arizona-3			
Moderate			
6 vs. 7 years	0.408	–0.708	.083
6 vs. 8 years	0.408	–1.417	.001*
7 vs. 8 years	0.408	–0.708	.083
Severe			
6 vs. 7 years	0.426	–1.000	.019
6 vs. 8 years	0.426	–1.727	< .001*
7 vs. 8 years	.0426	–0.727	.088
MWI			
Mild			
6 vs. 7 years	0.343	–0.647	.059
6 vs. 8 years	0.343	–1.294	< .001*
7 vs. 8 years	0.343	–0.647	.059
Moderate			
6 vs. 7 years	0.365	–0.667	.067
6 vs. 8 years	0.365	–1.133	.002*
7 vs. 8 years	0.365	–0.467	.201
PIU			
Mild			
6 vs. 7 years	0.343	–0.412	.230
6 vs. 8 years	0.343	–1.353	< .001*
7 vs. 8 years	0.343	–0.941	.006*

Note. Arizona-3 = Arizona Articulation Proficiency Scale–Third Edition; MWI = multiword intelligibility scores; PIU = percent intelligible utterances.

*$p < .0083$.

of speech sounds, it seems logical to expect that the lexical level would be impacted as well. It is surprising that PIU and parent ratings were not also (consistently) related to segmental integrity for these children. One explanation may be that the context provided in the PIU measure at ages 6 and 7 years allowed expert transcribers to overcome segmental and lexical limitations of the severely impaired speech signal.

Collectively, findings suggest that the interrelations among measures of speech production are influenced by severity of the child's speech impairment. Segmental integrity scores are related to intelligibility scores, particularly those at the lexical level for children with severe speech motor involvement, but not for children with mild or moderate involvement.

Differences Over Time Within Severity Group on Each Measure

We hypothesized that all variables would show change over the 2-year time frame of this study, in large part because of our previous research, which has shown that intelligibility is still growing in this time frame (Hustad, Sakash, Natzke, et al., 2019). However, this hypothesis did not hold true across all variables and severity groups.

Children in the Mild Group

Children in the mild group showed consistent growth on multiword intelligibility and PIU scores, but not on Arizona-3 scores. One variable that may be important to consider with regard to the findings of this longitudinal study is the age of the children. At the youngest age, children were 6 years old. Speech sound development is well underway and is approaching establishment of adult levels of integrity for most consonants at this age. For children in the mild group, Arizona-3 scores were very high, even at 6 years of age, reflecting developmentally appropriate segmental speech production skills that were nearing the ceiling in terms of development (see Table 3). Thus, there was little room for growth. Improvements in multiword intelligibility and PIU, even with unchanging Arizona-3 scores, suggest that children in the mild group were likely refining their speech motor control skills in ways that enhanced lexical and utterance-level intelligibility but were not detectable on standardized articulation tests focused at the level of the speech sound.

With regard to differences among the levels of measurement, several descriptive observations are notable. For children in the mild group, scores for PIU and multiword intelligibility were almost the same. These children also had very high Arizona-3 scores, suggesting that, when segmental integrity is strong, there may not be a differential effect in intelligibility scores based on level of measurement (utterance vs. word). In other words, contextual information is less important because the acoustic signal carries all necessary information to convey meaning.

Children in the Moderate Group

Children in the moderate group showed consistent growth on Arizona-3 and multiword intelligibility scores, indicating that they were making fine-grained improvements at the level of the speech sound and at the level of word production. Interestingly, these changes did not translate to coarser units, such as PIU and parent ratings of overall intelligibility to others. This finding is consistent with our correlational results, discussed previously, suggesting that measurements do not necessarily generalize or relate across levels.

For children in the moderate group, PIU was about 10 percentage points higher than multiword intelligibility. When the difference between utterance intelligibility and lexical intelligibility is considered from a conceptual view, this finding suggests that the contextual information that is available in language samples may provide a 10 percentage point advantage to children with moderate intelligibility deficits, again revealing a differential effect based on level of measurement.

Children in the Severe Group

In the severe group, children showed improvement only on Arizona-3 scores over time but did not show concomitant improvements in intelligibility as measured by multiword intelligibility or PIU. Thus, even though articulation got better, it did not improve functional speaking abilities nor did it improve parent ratings of their children. This finding is essentially the opposite of what we observed for children in the mild group who showed improved multiword intelligibility and PIU in the absence of change in Arizona-3 scores. Together, these findings highlight the complex interactions among variables ranging from the individual speech sound to the utterance (and associated suprasegmental and contextual variables that support the exchange of meaning).

For children in the severe group, PIU was about 35–45 percentage points higher than multiword intelligibility, so linguistic context provided by the interaction with a partner provided critical information that enhanced intelligibility. Examination of descriptive data (see Table 3) shows that Arizona-3 scores were surprisingly high for children in the severe group, in spite of multiword intelligibility scores that, on average, were below 50%, even at 8 years of age. This finding further highlights the complexity of speech performance and the potential disconnect between segmental level measures and measures of intelligibility in connected speech. Collectively, our findings suggest that the whole (connected speech) does not equal the sum of its parts (individual speech sounds).

A final finding of this study, cutting across severity groups, was that change in parent-reported measures of intelligibility to others was not significant for children in any group. These findings suggest that parent report for intelligibility should not be used to characterize change over time as it does not appear to be sensitive to change that has been quantified via other speech production measures. Although longitudinal studies of change in parent ratings of intelligibility over time have not been conducted prior to this study, cross-sectional research examining the effects of age on ICS ratings (McLeod et al., 2012) have suggested that improvements in intelligibility ratings tend to be very small (McLeod et al., 2015) and do not seem to show clear clinically meaningful change. Findings of this study also indicated that parent ratings were not correlated with other measures of speech production, suggesting that parent report measures may have captured something different from the other measures in this study. Previous research has suggested that parent perception of intelligibility was highly correlated with transcription intelligibility scores for children with CP (Hustad et al., 2012), but this work did not control for severity of speech deficits, and thus, results may have been related, in part, to severity as a moderating variable. Collectively, results of this study support the notion that parent ratings may provide valuable information regarding a broad view of a child's speech intelligibility, but that this information is not tied to direct measures that speech pathologists use in clinical assessment. Thus, we suggest that parent ratings, as measured in this study, should not be used instead of other direct measures of speech production, but rather in addition to those measures.

Limitations and Future Directions

This study has several important limitations. First, the participants in this study were grouped by severity according to single-word intelligibility scores. There are several ways that we could have grouped children. For example, we could have grouped children based on any of the other measures used in this study (i.e., multiword intelligibility, PIU, parent ratings, or Arizona-3 scores), and it is very likely that group composition would have been different, which, in turn, could have led to different findings. Therefore, it is important that results are considered in the context of these single-word intelligibility–based severity groupings. Future studies could seek to replicate this work using different grouping methodologies. Second, we examined a relatively small number of children. Although all children had CP and we attempted to reduce heterogeneity by grouping children according to speech intelligibility severity categories, the children in each group were still relatively heterogeneous with regard to their speech deficits. This heterogeneity is a feature of CP related to underlying neurological impairments, which are characteristically diverse, stemming from a considerable range of different neuropathologies. We did not examine a group of typically developing control children; therefore, we do not know if the findings from this study, particularly for the mild group, would be consistent with findings from children without CP who would be expected to be more homogeneous. In our earlier work, results have consistently shown that even those with CP who have no clinical evidence of speech motor involvement tend to have intelligibility deficits (Hustad, Sakash, Broman, & Rathouz, 2019; Hustad et al., 2012); thus, typically developing children may have a different pattern of results relative to findings from this study. Future research should examine interrelations among measures for typically developing children.

The extent to which findings from this study would generalize to other populations of children with speech intelligibility deficits is unknown but is an important area for future research. Studies should examine the interrelations among different measures of intelligibility and speech sound integrity for populations such as children with hearing impairment, children with speech sound disorders of unknown origin, children with cleft palate, and children with childhood apraxia of speech.

We examined differences on speech measures at 1-year intervals for three time points encompassing 2 years of growth. Not all measures of speech production examined in this study showed change over time between the ages of 6 and 8 years for all severity groups. Previous research on children with CP has suggested that growth in speech intelligibility is the steepest at earlier ages, particularly between the ages of 3 and 5 years, and that growth continues more slowly through 8 years of age (Hustad, Sakash, Natzke, et al., 2019). Future studies should examine a wider range of ages on measures that have not previously been examined in terms of their longitudinal growth over time in children with CP, for example, PIU, parent ratings of intelligibility, and Arizona-3 scores. This information would help us begin to understand rates and limits of change of different measures of speech for children with CP.

Clinical Implications

Results of this study clearly indicate that clinicians should carefully consider which assessment of speech performance is appropriate when evaluating functional speaking abilities and that measurement at multiple levels of performance is essential for comprehensively and systematically evaluating a child's ability to be understood across environments. For children with CP and mild-to-moderate speech impairment, findings from this study indicate a striking incongruence between the results of standardized articulation assessments and results obtained via other measures of speech production. These data, specifically for children with mild and moderate speech impairment, strongly indicate that assessments designed to inform SLPs about a child's proficiency in producing specific speech sounds should not lead to conclusions about speech intelligibility. Progress monitoring for these children should incorporate the judgments of unfamiliar listeners as to what a child has said in connected speech, as this measure was shown to be sensitive to change over time.

For children with CP and severe speech impairment, scores from a single-word articulation measure were correlated with multiword intelligibility and also were found to be sensitive to change over time. However, clinicians should remember that articulation assessments provide little information as to functional speaking performance in real-world situations. It is important to highlight that it is this group of children who demonstrated the greatest gap in performance between single-word tasks and conversational tasks, and therefore, it is this group of children who benefit the most when listeners are able to leverage contextual cues to aid their interpretation of the speech signal. It is critical that clinicians serving children with severe speech motor impairment consider how to provide additional supports, such as topic cues, to help set the stage for a successful dialogue. Using unfamiliar listeners to transcribe the connected speech of these children, with and without contextual cues, may be an effective way to monitor progress and continually assess how to increase participation across environments.

We examined children with CP, a population with considerable heterogeneity in their speech production abilities. Notably, speech deficits for many children with CP are neurologically based and reflect deficits across multiple speech subsystems. The extent to which our findings may generalize to other populations of children, such as those with speech sound disorders or childhood apraxia of speech, is unknown. However, it would not be surprising if similar findings were observed for children with speech disorders that were more severe in nature, particularly those with more complex or neurologically based underlying impairments. If such findings do hold for other populations of children with reduced intelligibility, our results would have important implications for assessment of speech performance.

Acknowledgments

This study was funded by National Institute on Deafness and Other Communication Disorders Grant R01DC009411 (awarded to Katherine Hustad). Support was also provided by a core grant to the Waisman Center (U54 HD090256) from the National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank the children and their families who participated in this research, as well as the graduate and undergraduate students in the Wisconsin Intelligibility, Speech, and Communication Lab of the University of Wisconsin–Madison who assisted in data collection and data reduction.

References

Allison, K. M., & Hustad, K. C. (2018). Acoustic predictors of pediatric dysarthria in cerebral palsy. *Journal of Speech, Language, and Hearing Research, 61*(3), 462–478. https://doi.org/10.1044/2017_JSLHR-S-16-0414

Bickham, R. S., Ranganathan, K., Wombacher, N. R., Shapiro, D. N., Carlozzi, N. E., Baker, M. K., Vercler, C. J., Warschausky, S. A., Buchman, S. R., & Waljee, J. F. (2017). Speech perceptions and health-related quality of life among children with cleft lip and palate. *Journal of Craniofacial Surgery, 28*(5), 1264–1268. https://doi.org/10.1097/SCS.0000000000003718

Binger, C., Ragsdale, J., & Bustos, A. (2016). Language sampling for preschoolers with severe speech impairments. *American Journal of Speech-Language Pathology, 25*(4), 493–507. https://doi.org/10.1044/2016_AJSLP-15-0100

Connolly, J. H. (1986). Intelligibility: A linguistic view. *International Journal of Language & Communication Disorders, 21*(3), 371–376. https://doi.org/10.3109/13682828609019848

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*(2), 246–269. https://doi.org/10.1044/jshr.1202.246

Dickinson, H. O., Parkinson, K. N., Ravens-Sieberer, U., Schirripa, G., Thyen, U., Arnaud, C., Beckung, E., Fauconnier, J., McManus, V., Michelsen, S. I., Parkes, J., & Colver, A. F. (2007). Self-reported quality of life of 8–12 year old children with cerebral palsy: A cross-sectional European study. *The Lancet, 369*(9580), 2171–2178. https://doi.org/10.1016/S0140-6736(07)61013-7

Ertmer, D. J. (2010). Relationships between speech intelligibility and word articulation scores in children with hearing loss. *Journal of Speech, Language, and Hearing Research, 53*(5), 1075–1086. https://doi.org/10.1044/1092-4388(2010/09-0250)

Ertmer, D. J., & Goffman, L. A. (2011). Speech production accuracy and variability in young cochlear implant recipients: Comparisons with typically developing age-peers. *Journal of Speech, Language, and Hearing Research, 54*(1), 177–189. https://doi.org/10.1044/1092-4388(2010/09-0165)

Fox, C. M., & Boliek, C. A. (2012). Intensive voice treatment (LSVT LOUD) for children with spastic cerebral palsy and dysarthria. *Journal of Speech, Language, and Hearing Research, 55*(3), 930–945. https://doi.org/10.1044/1092-4388(2011/10-0235)

Fudala, J. B. (2001). *Arizona Articulation Proficiency Scale–Third Edition*. Western Psychological Services.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various coefficients of interrater reliability and agreement*. https://CRAN.R-project.org/package=irr

Goldman, R., & Fristoe, M. (2015). *Goldman-Fristoe Test of Articulation–Third Edition*. PsychCorp.

Heilmann, J., Nockerts, A., & Miller, J. F. (2010). Language sampling: Does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools, 41*(4), 393–404. https://doi.org/10.1044/0161-1461(2009/09-0023)

Hodge, M., & Daniels, J. (2007). *TOCS+ intelligibility measures*. University of Alberta.

Hodge, M., & Gotzke, C. L. (2014). Criterion-related validity of the Test of Children's Speech sentence intelligibility measure for children with cerebral palsy and dysarthria. *International Journal of Speech-Language Pathology, 16*(4), 417–426. https://doi.org/10.3109/17549507.2014.930174

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*(3), 167–175. https://doi.org/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*(6), 1496–1513. https://doi.org/10.1044/1092-4388(2010/09-0176)

Hustad, K. C., Oakes, A., & Allison, K. (2015). Variability and diagnostic accuracy of speech intelligibility scores in children. *Journal of Speech, Language, and Hearing Research, 58*(6), 1695–1707. https://doi.org/10.1044/2015_JSLHR-S-14-0365

Hustad, K. C., Sakash, A., Broman, A. T., & Rathouz, P. J. (2019). Differentiating typical from atypical speech production in 5-year-old children with cerebral palsy: A comparative analysis. *American Journal of Speech Language Pathology, 28*(2S), 807–817. https://doi.org/10.1044/2018_AJSLP-MSC18-18-0108

Hustad, K. C., Sakash, A., Natzke, P. E. M., Broman, A. T., & Rathouz, P. J. (2019). Longitudinal growth in single word intelligibility among children with cerebral palsy from 24 to 96 months of age: Predicting later outcomes from early speech production. *Journal of Speech, Language, and Hearing Research, 62*(6), 1599–1613. https://doi.org/10.1044/2018_JSLHR-S-18-0319

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*(4), 1177–1189. https://doi.org/10.1044/1092-4388(2011/11-0083)

Kent, R. D. (1993). Speech intelligibility and communicative competence in children. In A. P. Kaiser & D. B. Gray (Eds.), *Enhancing children's communication: Foundations for intervention* (Vol. 2, pp. 223–239). Brookes.

Kent, R. D., Miolo, G., & Bloedel, S. (1994). The intelligibility of children's speech: A review of evaluation procedures. *American Journal of Speech-Language Pathology, 3*(2), 81–95. https://doi.org/10.1044/1058-0360.0302.81

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54*(4), 482–499. https://doi.org/10.1044/jshd.5404.482

Lagerberg, T. B., Åsberg, J., Hartelius, L., & Persson, C. (2014). Assessment of intelligibility using children's spontaneous speech: Methodological aspects. *International Journal of Language & Communication Disorders, 49*(2), 228–239. https://doi.org/10.1111/1460-6984.12067

Landa, S., Pennington, L., Miller, N., Robson, S., Thompson, V., & Steen, N. (2014). Association between objective measurement of the speech intelligibility of young people with dysarthria and listener ratings of ease of understanding. *International Journal of Speech-Language Pathology, 16*(4), 408–416. https://doi.org/10.3109/17549507.2014.927922

Lee, J., Hustad, K. C., & Weismer, G. (2014). Predicting speech intelligibility with a multiple speech subsystems approach in children with cerebral palsy. *Journal of Speech, Language, and Hearing Research, 57*(5), 1666–1678. https://doi.org/10.1044/2014_JSLHR-S-13-0292

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. NCTE Committee on Research Report No. 18.

Malmenholt, A., McAllister, A., & Lohmander, A. (2019). Orofacial function, articulation proficiency, and intelligibility in 5-year-old children born with cleft lip and palate. *The Cleft Palate–Craniofacial Journal, 56*(3), 321–330. https://doi.org/10.1177/1055665618783154

McLeod, S., Crowe, K., & Shahaeian, A. (2015). Intelligibility in Context Scale: Normative and validation data for English-speaking preschoolers. *Language, Speech, and Hearing Services in Schools, 46*(3), 266–276. https://doi.org/10.1044/2015_LSHSS-14-0120

McLeod, S., Harrison, L. J., & McCormack, J. (2012). The Intelligibility in Context Scale: Validity and reliability of a subjective rating measure. *Journal of Speech, Language, and Hearing Research, 55*(2), 648–656. https://doi.org/10.1044/1092-4388(2011/10-0130)

Miller, J. F., Andriacchi, K., & Nockerts, A. (2011). *Assessing language production using SALT software: A clinician's guide to language sample analysis*. SALT Software.

Namasivayam, A. K., Pukonen, M., Goshulak, D., Yu, V. Y., Kadis, D. S., Kroll, R., Pang, E. W., & De Nil, L. F. (2013). Relationship between speech motor control and speech intelligibility in children with speech sound disorders. *Journal of Communication Disorders, 46*(3), 264–280. https://doi.org/10.1016/j.jcomdis.2013.02.003

Nordberg, A., Miniscalco, C., Lohmander, A., & Himmelmann, K. (2013). Speech problems affect more than one in two children with cerebral palsy: Swedish population-based study. *Acta Paediatrica, 102*(2), 161–166. https://doi.org/10.1111/apa.12076

Palisano, R. J., Rosenbaum, P., Bartlett, D., & Livingston, M. H. (2008). Content validity of the expanded and revised Gross Motor Function Classification System. *Developmental Medicine & Child Neurology, 50*(10), 744–750. https://doi.org/10.1111/j.1469-8749.2008.03089.x

Pennington, L., Virella, D., Mjøen, T., da Graça Andradad, M., Murray, J., Colver, A., Himmelmann, K., Rackauskaite, G., Greitane, A., Prasauskiene, A., Andersen, G., & de la Cruz, J. (2013). Development of the Viking Speech Scale to classify the speech of children with cerebral palsy. *Research in Developmental Disabilities, 34*(10), 3202–3210. https://doi.org/10.1016/j.ridd.2013.06.035

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., 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*(2), 333–349. https://doi.org/10.1044/1092-4388(2009/08-0183)

Weismer, G., & Martin, R. E. (1992). Acoustic and perceptual approaches to the study of intelligibility. In R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement and management* (pp. 67–118). John Benjamins. https://doi.org/10.1075/sspcl.1.04wei

Whitehill, T. L. (2002). Assessing intelligibility in speakers with cleft palate: A critical review of the literature. *The Cleft Palate–Craniofacial Journal, 39*(1), 50–58. https://doi.org/10.1597/1545-1569(2002)039<0050:AIISWC>2.0.CO;2

Williams, A. L., McLeod, S., & McCauley, R. J. (2010). *Interventions for speech sound disorders in children*. Brookes.

Yoder, P. J., Woynaroski, T., & Camarata, S. (2016). Measuring speech comprehensibility in students with Down syndrome. *Journal of Speech, Language, and Hearing Research, 59*(3), 460–467. https://doi.org/10.1044/2015_JSLHR-S-15-0149

Yorkston, K. M., & Beukelman, D. R. (1980). A clinician-judged technique for quantifying dysarthric speech based on single-word intelligibility. *Journal of Communication Disorders, 13*(1), 15–31. https://doi.org/10.1016/0021-9924(80)90018-0

Yorkston, K. M., Beukelman, D. R., Strand, E. A., & Bell, K. (1999). *Management of motor speech disorders in children and adults* (2nd ed.). Pro-Ed.

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*(1), 55–66. https://doi.org/10.1044/1058-0360.0501.55