JSLHR

Research Article

Variability and Diagnostic Accuracy of Speech Intelligibility Scores in Children

Katherine C. Hustad,[a,b] Ashley Oakes,[b] and Kristen Allison[a,b]

Purpose: We examined variability of speech intelligibility scores and how well intelligibility scores predicted group membership among 5-year-old children with speech motor impairment (SMI) secondary to cerebral palsy and an age-matched group of typically developing (TD) children.
Method: Speech samples varying in length from 1–4 words were elicited from 24 children with cerebral palsy (mean age 60.50 months) and 20 TD children (mean age 60.33 months). Two hundred twenty adult listeners made orthographic transcriptions of speech samples (*n* = 5 per child).
Results: Variability associated with listeners made a significant contribution to explaining the variance in intelligibility scores for TD and SMI children, but the magnitude was greater for TD children. Intelligibility scores differentiated very well between children who have SMI and TD children when intelligibility was at or below approximately 75% and above approximately 85%.
Conclusions: Intelligibility seems to be a useful clinical tool for differentiating between TD children and children with SMI at 5 years of age; however, there is considerable variability within and between listeners, highlighting the need for more than one listener per child to ensure validity of an intelligibility measure.

Intelligible speech is critical for successful spoken communication. Intelligibility has been defined as the extent to which an acoustic signal, generated by a speaker, can be correctly recovered by a listener (Kent, Weismer, Kent, & Rosenbek, 1989). Central to this definition is the dyadic nature of intelligibility, requiring at a minimum a speaker who produces a speech signal and a listener who receives the signal. Thus, intelligibility is a product of the joint efforts of the speaker and the listener (Lindblom, 1990). Intelligibility has received considerable attention in the adult dysarthria literature (e.g., Borrie, McAuliffe, & Liss, 2012; D'Innocenzo, Tjaden, & Greenman, 2006; Kim, Hasegawa-Johnson, & Perlman, 2011; Liss, Spitzer, Caviness, & Adler, 2002; McAuliffe, Carpenter, & Moran, 2010; Rong, Loucks, Kim, & Hasegawa-Johnson, 2012). Many variables have been shown to influence intelligibility, highlighting the complexity of the construct. These variables can be separated into those associated with the speaker, the listener, and contextual factors (Hustad & Weismer, 2007). Historically, research has focused on understanding how characteristics of the signal relate to intelligibility (see Kim et al., 2011; Turner, Tjaden, & Weismer, 1995; Weismer, Jeng, Laures, Kent, & Kent, 2001; Yunusova, Weismer, Kent, & Rusche, 2005). In recent years, efforts have focused on perceptual variables related to the listener in studies of intelligibility, including studies of perceptual learning in listeners when presented with dysarthric speech, as well as listening strategies and mechanisms (Borrie, McAuliffe, & Liss, 2012; Borrie, McAuliffe, Liss, et al., 2012; Borrie, McAuliffe, Liss, O'Beirne, & Anderson, 2013; Choe, Liss, Azuma, & Mathy, 2012; Kim & Nanney, 2014). The importance of the listener is now widely recognized in intelligibility research.

Intelligibility in Children

In children, acquisition of intelligible speech is a protracted developmental process, beginning early in the first year of life with vocal play, babbling, and word approximations and continuing through childhood, culminating in the ability to produce fully intelligible, adultlike connected speech. Considerable attention has been devoted to the study of acquisition of speech sounds at a segmental level. Collectively, research indicates that by about 5 to 6 years of age, English-speaking children produce most speech sounds accurately at least some of the time, but they are not expected to have mastered all sounds until about 8 years of age (Sander, 1972; Smit, Hand, Freilinger, Bernthal, & Bird, 1990). However, the way in which segmental acquisition

[a]University of Wisconsin–Madison
[b]Waisman Center, University of Wisconsin–Madison
Correspondence to Katherine C. Hustad: kchustad@wisc.edu

Editor: Jody Kreiman
Associate Editor: Julie Liss
Received December 29, 2014
Revision received June 30, 2015
Accepted August 17, 2015
DOI: 10.1044/2015_JSLHR-S-14-0365

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

translates to intelligible speech at the level of words and sentences is not well understood. Studies of phonetic transcription of articulation errors have been used in an effort to examine how specific segmental errors associated with different kinds of speech disorders (e.g., dysarthria, phonological disorders, cleft palate speech, and hearing-impaired speech) may predict intelligibility deficits. Results have varied considerably, and this variation does not appear related to etiology of the speech disorder (Weismer, 2008).

Listeners are often able to deduce lexical information correctly even in the presence of speech sound errors, highlighting the importance of studying intelligibility as an entity separate from segmental acquisition in children. However, growth trajectories for acquisition of intelligible speech, the particular age at which children are expected to become fully intelligible to everyday listeners, and the range of acceptable variability in intelligibility among children by age have not been clearly established with empirical data. A limited number of studies suggest that children should be fully intelligible to unfamiliar listeners by about 4 years of age (Coplan & Gleason, 1988; Flipsen, 2006; Weiss, 1982). Other studies using standard orthographic transcription of words and sentences by unfamiliar listeners provide evidence that 4-year-old children are not fully intelligible (M. M. Hodge & Gotzke, 2014a; Hustad, Schueler, Schulz, & DuHadway, 2012). The methodologies used across studies have varied considerably, making comparisons difficult. For example, existing studies examining children have primarily used listeners who are "experts" (commonly speech-language pathologists, phoneticians, or graduate students in speech-language pathology). In addition, studies have examined intelligibility at different linguistic levels (i.e., single words, sentences, and spontaneous speech), and this appears to be an important variable. The lack of well-established normative data makes it challenging to identify age-appropriate expectations that can serve as benchmarks for typical versus atypical intelligibility.

Reduced intelligibility is associated with a wide range of disorders in children, including cleft lip and palate, speech sound disorders, hearing impairment, dysarthria, and childhood apraxia of speech. Although the underlying causes and specific speech characteristics vary greatly across populations, persistent intelligibility deficits can have important repercussions, including social and educational participation and reduced quality of life (Dickinson et al., 2007; Schneider, Gurucharri, Gutierrez, & Gaebler-Spira, 2001). Enhancing intelligibility is often a key objective of speech therapy (M. M. Hodge & Gotzke, 2014a; Landa et al., 2014).

In this study, intelligibility of children with cerebral palsy (CP) is of particular interest. The estimated 2.5 per 1,000 children who have CP (Paneth, Hong, & Korzeniewski, 2006) are at significant risk for the motor speech disorder dysarthria. Our recent work suggests that at 4 years of age, up to 75% of children with CP may have speech motor impairment (SMI; Hustad, Gorton, & Lee, 2010). Dysarthria can range from very mild to severe and profound. It can be difficult to definitively identify dysarthria in young children with CP for at least four reasons: (a) The onset of talking may be significantly delayed (Hustad, Allison, McFadd, & Riehle, 2013); thus, subsequent speech development may also be delayed. (b) Symptoms may be mild or ambiguous. (c) There is a wide range of variability associated with typical speech development that makes it difficult to separate typical variability from atypical. (d) There is overlap between potential dysarthric speech characteristics and typical developmental error patterns. A standard set of expectations for intelligibility development by age could serve as an important diagnostic tool for differentiating among children whose intelligibility is within the normal range from those whose intelligibility is not. This information would be useful for children with a wide range of etiologies who may experience reductions in intelligibility.

Because so many variables affect intelligibility, no one measure can comprehensively characterize it. Nonetheless, measures of speech intelligibility have been used for a variety of clinical purposes, such as providing an index of functional ability (Yorkston, Beukelman, Strand, & Bell, 1999), a measure of severity of the speech disorder (Weismer, 2008; Weismer & Martin, 1992), and as a basis of comparison for documenting and monitoring change in speech performance (Yorkston et al., 1999). Important clinical decisions are often made on the basis of speech intelligibility data, highlighting the necessity for a thorough and quantitative understanding of the variables that contribute to intelligibility. Given the clinical utility of intelligibility measures, one important question is whether intelligibility scores can be used to differentiate children with dysarthria from typically developing (TD) peers, and if so, what are the specific cutpoints for typical and atypical intelligibility expectations at a given age?

Listener Variability in Intelligibility Measurement

One observation that seems to be common among studies of intelligibility involving both children and adults across disorder populations is that there is considerable variability in the performance of different listeners who hear the same speaker (interlistener variability; see Hustad & Cahill, 2003; Hustad et al., 2012; Lam & Tjaden, 2013; Montag, AuBuchon, Pisoni, & Kronenberger, 2014; Pennington et al., 2013). In fact, studies report standard deviations within a given speaker across listeners of up to 17% for individuals with the most severe dysarthria and slightly lower for those speakers with less severe dysarthria (Hustad & Cahill, 2003). The presence of this variability suggests that more than one listener is necessary to obtain an intelligibility score that is ecologically valid. However, the extent to which interlistener variability affects intelligibility measurement has not been systematically investigated or quantified.

Another source of potential variability lies within individual listeners. Within-listener performance can be considered in at least two different ways. Test–retest reliability examines consistency of performance when measurement is repeated for the same listener, the same speaker, and the same speech stimuli. Test–retest reliability measures compare performance on the first and second administrations of a test and provide important information on stability. A second related type of within-listener variability is the variance associated with a mean intelligibility score for an individual listener, hereafter referred to as *intralistener variability*. Intralistener variability has not been reported in previous intelligibility studies, but research has shown that stimulus characteristics, such as length of utterance, can have an important impact, so it is reasonable to expect intralistener variability to differ on the basis of the nature of the speech stimuli. To our knowledge, there have not been studies of intelligibility in children (disordered or typical) or studies of intelligibility in adults that have directly addressed either of these two types of reliability. Note that studies of adults with dysarthria suggest that listeners can learn as a listening task progresses (Hustad & Cahill, 2003), which may have an important impact on intelligibility measurement. Both intralistener variability and within-listener test–retest reliability are important avenues of investigation given that intelligibility measurement hinges on listeners as much as it does on the speaker.

Systematic study of the extent to which interlistener variability, intralistener variability, and test–retest reliability may affect intelligibility is an important step in understanding the range of variables that account for intelligibility, as the listener is a necessary element of any intelligibility measure. In particular, it is important to understand the stability of listener behavior and how it might be influenced by children's speech at different ages and by different production abilities.

In the present study, we sought to (a) examine variability of speech intelligibility scores and to (b) determine how well intelligibility scores differentiate children who have dysarthria secondary to CP, hereafter referred to as *children with speech motor impairment* (SMI), following our previously established classification terminology (Hustad et al., 2010, 2012) and children who are TD. Within these two broad sets of questions, we addressed the following specific questions:

a1) Are there differences between intralistener and interlistener variability for children with SMI and for TD children?

a2) Are there differences between SMI versus TD groups on intra- and interlistener variability in intelligibility scores?

a3) To what extent do intra- and interlistener variability account for the variance in intelligibility scores within SMI and TD children?

a4) How good is test–retest reliability for listeners of SMI and TD children?

b1) How accurate are intelligibility scores globally in differentiating between and classifying TD and SMI children?

b2) Are there specific cutpoints in which intelligibility scores perform better at differentiating between groups of children?

Method

Participants

Children with CP

Twenty-four children with CP participated as speakers. All children were participants in a longitudinal study on communication development. Inclusion criteria for the larger study required that children (a) have a medical diagnosis of CP; and (b) have hearing abilities within normal limits, as documented by either formal audiological evaluation or distortion product otoacoustic emission screening. For the present study, we selected children from the larger cohort who met the following additional criteria: (a) presence of SMI, as determined by clinical judgment of two speech pathologists with expertise in CP; and (b) ability to produce utterances that were four words in length.

The mean age across children was 60.50 months (*SD* = 5.97). The sample included nine boys (mean age 60.78 months [*SD* = 6.00]) and 15 girls (mean age 60.33 months [*SD* = 6.16]). Table 1 presents demographic characteristics of children, including medical diagnoses, receptive language skills, and Gross Motor Function Classification System (Palisano et al., 1997) rating.

TD Children

Twenty TD children also participated as speakers. Children were recruited from the local community, including from a university preschool, through word of mouth, and through public postings. Inclusion criteria required that children (a) have TD speech, (b) have TD language, (c) have no history of developmental delay per parent report, and (d) have hearing abilities within normal limits, as documented by distortion product otoacoustic emission screening. Standardized speech and language screening measures and audiologic screening were administered to ensure that children met inclusion criteria. Speech was screened using the Arizona Articulatory Proficiency Scale–Third Edition (Fudala, 2001). Language was screened using the Preschool Language Scale Screening Test–Fourth Edition (Zimmerman, Steiner, & Pond, 2002). The mean age across TD children who were included in the study was 58.65 months (*SD* = 4.90). The sample included eight boys (mean age 56.49 months [*SD* = 3.12]) and 12 girls (mean age 60.10 months [*SD* = 5.46]).

Nondisabled Adult Listeners

Two hundred twenty nondisabled adults participated as listeners. Listeners were recruited from the university community via public postings and were primarily undergraduate students. Listeners were compensated monetarily for their participation. Five different listeners were randomly assigned to each child; each listener heard only one child producing all stimulus material. Inclusion criteria required that listeners (a) pass pure tone hearing screening administered via headphones at 25 dB HL for 250, 500, 1k, 4k, and 6 kHz bilaterally; (b) be between 18 and 45 years of age;

Table 1. Demographic characteristics of children with cerebral palsy and speech motor impairment.

Child	Age	Sex	GMFCS rating[a]	Diagnosis	Language comprehension SS[b]
1	65.0	M	4	Quadriplegia	70
2	62.0	F	4	Diplegia	74
3	66.0	F	2	Diplegia	76
4	71.0	F	5	Quadriplegia	76
5	67.0	F	1	Diplegia	83
6	62.0	F	2	Hemiplegia (left)	85
7	62.0	F	2	Hemiplegia (right)	87
8	66.0	M	2	Hemiplegia (right)	87
9	52.0	M	4	Quadriplegia	91
10	50.0	F	3	Quadriplegia	94
11	48.0	F	2	Hemiplegia (right)	96
12	53.0	M	1	Ataxic	98
13	68.0	M	4	Hemiplegia (right)	98
14	61.0	M	4	ataxic	98
15	62.0	M	1	Diplegia	100
19	61.0	F	1	Hemiplegia (right)	102
17	65.0	M	2	Hemiplegia (right)	104
18	62.0	F	4	Quadriplegia	106
19	55.0	F	1	Hemiplegia (left)	111
20	55.0	F	2	Unknown	113
21	55.0	M	1	Hypotonic	113
22	64.0	F	2	Triplegia	119
23	60.0	F	4	Diplegia	124
24	60.0	F	2	Dyskinetic	128

Note. GMFCS = Gross Motor Function Classification System; SS= standard scores; M= male; F = female.

[a]GMFCS ratings reflect clinician judgment of gross motor classification system level (Palisano et al., 1997). [b]Standard language scores were obtained from the Test for Auditory Comprehension of Language–Third Edition (Carrow-Woolfolk, 1999).

(c) have no more than incidental experience listening to or communicating with persons having communication disorders; (d) be a native speaker of American English; and (e) have no identified language, learning, or cognitive disabilities per self-report. Listeners were 74 males and 146 females. No effort was made to balance the sex distribution of listeners, as this was not a variable of interest. The mean age of listeners was 21.50 years (*SD* = 4.2).

Acquisition of Speech Samples From Children: Materials and Procedure

Children produced a standard corpus of speech stimuli, which were recorded and played back to listeners who made intelligibility judgments. Speech stimuli produced by children were taken from the Test of Children's Speech (TOCS+; Hodge & Daniels, 2007). The TOCS+ is a set of single words and sentences that systematically vary in length and are developmentally appropriate (lexically, phonetically, syntactically, and morphologically) for children. In this study, we used the same stimuli for each child to ensure equivalence among utterances and children. Children produced 38 different single-word utterances and 30 different multiword utterances. Multiword utterances were evenly divided among two-, three-, and four-word lengths. We did not include utterances of longer lengths because many children with CP and SMI were unable to produce a complete corpus of utterances above a four-word level. In addition, for the age of participants in this study, the length and linguistic level of stimuli up to four words were well within speech and language production abilities and thus were considered to be ecologically valid.

The research protocol was administered by a speech-language pathologist in a sound-attenuating room. Speech samples from children were recorded using a digital audio recorder (Marantz PMD 570, D & M Holdings Inc., Tokyo, Japan) at a 44.1-kHz sampling rate (16-bit quantization). A condenser studio microphone (Audio-Technica AT4040, Audio-Technica U.S., Inc., Stow, OH) was positioned next to each child using a floor stand and was located approximately 18 in. from the child's mouth. The level of the signal was monitored and adjusted on a mixer (Mackie 1202 VLZ, Mackie Designs Inc., Woodinville, WA) to obtain optimized recordings and to avoid peak clipping.

Speech samples were obtained from children by using imitation of the target corpus of words and sentences. This paradigm allowed us to compare listener transcription data with known target responses, thus ensuring that intelligibility scores were an accurate reflection of whether listeners perceived the target words correctly. To ensure consistency across modeled productions, adult recordings of each target utterance were presented via a laptop computer to each child along with an image depicting the utterance. Children then repeated what they heard following the recorded model. Productions were monitored online by a research assistant to ensure that samples were free of overlap with examiner speech and free of extraneous noises. Children were asked to repeat utterances when these criteria were not met. All children were able to produce all speech stimuli. The full research protocol took approximately 2 hr, and all children tolerated this without difficulty.

Acquisition of Intelligibility Data: Materials and Procedure

Digital audio recordings were transferred to a personal computer. Recordings were edited to remove extraneous noises and the examiner's voice; individual files were created for each stimulus utterance produced by each child. Audio samples were peak amplitude normalized to assure that maximum loudness levels of the recorded speech samples were the same across children and utterances, while preserving the amplitude contours of the original productions. This also enabled calibration to peak output levels for playback to listeners.

Listeners completed two listening tasks: in one, they heard a single child producing all single-word stimulus utterances, and in the other, they heard the same child producing all multiword stimulus utterances. The order of presentation for the single-word listening task and the multiword listening task was counterbalanced among the listeners of each child to prevent a potential order effect. The individual stimulus items within each task were randomized for each listener so that no two listeners heard the stimulus items in the same sequence. In addition, 15% of stimulus items for each of the two tasks were randomly selected and presented twice to each listener to examine intralistener reliability. Stimuli included six single-word utterances and five sentences. The average separation of first and second presentation of the same items across listeners was 18 stimulus items.

During the experiment, listeners were seated individually in a sound-attenuating suite in front of a 19-in. flat panel computer screen with a keyboard placed directly in front of them. An external speaker, connected to a computer, was situated directly beneath the computer screen. The peak audio output level was calibrated to approximately 75 dB SPL from where listeners were seated and was checked periodically to ensure that all listeners heard stimuli at the same output level.

Speech stimuli were delivered via an in-house computer program that presented stimulus utterances and stored typed orthographic transcriptions. Listeners were allowed to hear each utterance one time. Listeners were told that the purpose of the study was to determine how understandable children were to unfamiliar listeners like themselves. They were instructed that children would be producing real words and to take their best guess if they were unsure as to what the child said. Listeners were provided with instructions on how to use the experimental software to advance through the experiment. In addition, they heard two multiword and two single-word utterances as examples to familiarize themselves with the experimental task. Data from the example utterances were excluded from analyses. Data collection from listeners took approximately 30–45 min.

Analysis of Data: Speech Intelligibility

Orthographic Transcription of Intelligibility

Orthographic transcriptions of speech samples for each child were generated by five independent listeners per child. Transcriptions were scored by using our in-house computer program. The program automatically tallied the number of transcribed words that were an exact phonemic match to the stimulus words in the sentences produced by the children. Misspellings and homonyms were accepted as correct, as long as all phonemes in the spoken version of the transcribed words matched the target words. Dependent measures and their operational definitions and calculations were as follows: (a) The mean intelligibility for each listener was obtained by calculating the number of words identified correctly within utterances of each length (one to four words), resulting in four separate intelligibility scores. The mean across utterance lengths was then calculated to yield an intelligibility score for each listener. (b) Intralistener variability for each listener was calculated as the standard deviation of the means for utterances of each length within each listener. Thus, intralistener variability captured the joint impact of stimulus length and variance within a given listener. This was important because stimulus length is known to be a variable that affects intelligibility (Allison & Hustad, 2014; Hustad et al., 2012). (c) Mean intelligibility for each child was obtained by averaging mean intelligibility scores derived for each listener across the five listeners who heard each child. (d) Interlistener variability for each child was obtained by calculating standard deviations of mean intelligibility scores across the five listeners per child within utterances of each length, resulting in four different standard deviations for each child. These values were then averaged to capture an overall measure of interlistener variability across utterance lengths for each child. In this way, our measure of interlistener variability also captured the impact of stimulus length across listeners. Thus, our measures of intra- and interlistener variability were conceptually and psychometrically comparable. (e) Intralistener reliability was obtained by comparing intelligibility scores for the first and second presentation of the randomly selected subset of stimuli that were played twice for each child and listener.

Statistical Analyses

To examine our first set of questions pertaining to variability of speech intelligibility scores, we completed four sets of analyses: (a) Differences between intralistener versus interlistener variability in intelligibility scores for SMI and for TD groups were tested by using a planned comparisons approach with two pairwise *t* tests and an alpha level of .01 that was partitioned evenly between the tests. Probability levels less than or equal to .005 were necessary for significance. (b) Differences between SMI versus TD groups for intra- and interlistener variability in intelligibility scores were also examined with planned comparisons by using two pairwise *t* tests and an alpha level of .01 that was partitioned evenly between the tests. Probability levels less than or equal to .005 were necessary for significance. (c) The extent to which intra- and interlistener variability account for variance in intelligibility scores within SMI and TD groups was examined by using a regression model with interlistener variability and intralistener variability simultaneously regressed onto intelligibility scores for children in each group. We used an alpha level of .01 for each of the two regression analyses. Patterns of findings between the two groups were compared descriptively. (d) Intralistener reliability was examined by comparing differences in mean intelligibility scores for the first and second presentations of a subset of speech stimuli using a planned comparisons with pairwise *t* tests. We used an alpha level of .01 that was partitioned evenly among the two inferential tests; thus, probability levels less than or equal to .005 were necessary for significance. We also computed Cohen's kappa statistics, examining item-by-item consistency of performance for the first and second presentations of stimuli within each child. We obtained the mean kappa across children within SMI and TD groups and report the descriptive findings pooled by group.

To examine how well intelligibility scores differentiated between children with SMI and TD children, we completed two sets of analyses: (a) We computed a receiver operating characteristic (ROC) curve and determined the area under the curve as a global measure of test accuracy. From the coordinates of the ROC curve, we calculated sensitivity, specificity, and positive and negative likelihood ratios for intelligibility scores at 5% intervals. (b) To determine whether there were specific cutpoints in which intelligibility scores performed better at differentiating between groups of children, we used binary logistic regression to classify children into TD and SMI groups on the basis of intelligibility scores.

Results

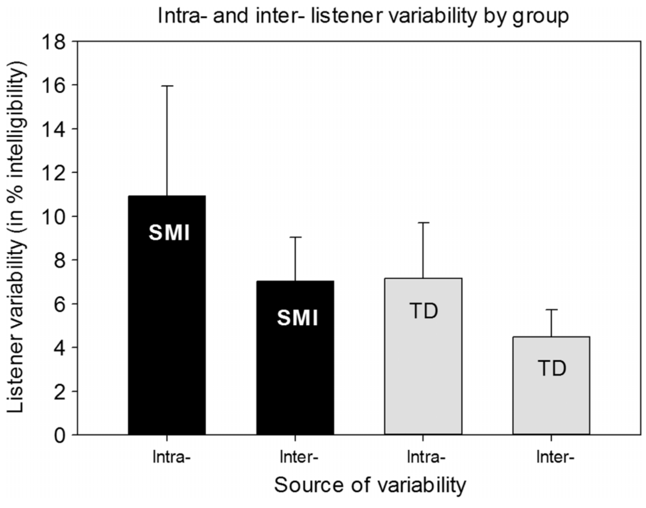
Question 1: Variability of Intelligibility Scores

Results of the four sets of analyses are presented by specific question.

(a) Are there differences between intralistener versus interlistener variability for children with SMI and for TD children? Descriptive data are shown in Figure 1. Descriptive results suggested that intralistener variability was greater than interlistener variability within both groups of children. Inferential statistics were consistent with descriptive observations. Intralistener variability was significantly greater than interlistener variability for children in the SMI group (mean difference = 3.87, 95% confidence interval [CI] [3.00, 4.76]), $t(119) = 8.74$, $p < .001$, $d = .79$; and for children in the TD group (mean difference = 2.67, 95% CI [2.12, 3.21]), $t(99) = 9.73$, $p < .001$, $d = .79$.

(b) Are there differences between SMI versus TD groups on intra- and interlistener variability in intelligibility scores? Descriptive data are shown in Figure 1 and suggest that children with SMI had higher intralistener variability, as well as higher interlistener variability than TD children. The group difference was significant for intralistener (mean difference = 3.76, 95% CI [2.72, 4.80]), $t(218) = 6.78$, $p < .001$, $d = .94$; and for interlistener variability (mean difference = 2.55, 95% CI [2.11, 2.98]), $t(218) = 11.03$, $p < .001$, $d = 1.52$.

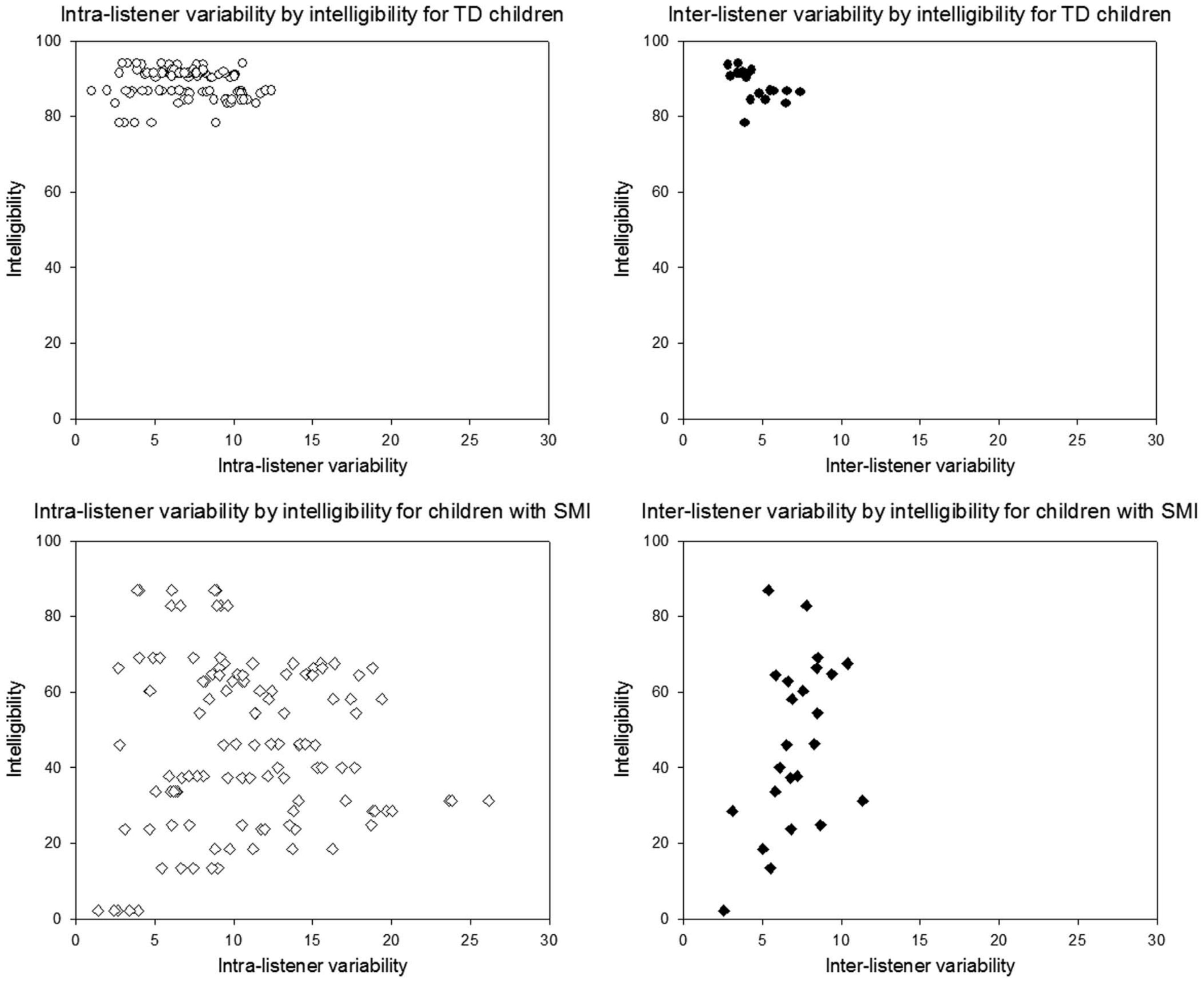
Figure 1. Mean intra- and interlistener variability by group. SMI = speech motor impairment; TD = typically developing.



(c) To what extent do intra- and interlistener variability account for the variance in intelligibility scores within SMI and TD children? Figure 2 shows scatterplots of intra- and interlistener variability relative to intelligibility scores for each group of children. Results of regression analyses indicated that when both intra- and interlistener variability were simultaneously regressed onto intelligibility, a significant amount of variance in intelligibility scores was accounted for by the model (see Table 2). This was true for children in the SMI group, $F(2, 117) = 12.812$; $p < .001$, and for children in the TD group, $F(2, 97) = 20.378$; $p < .001$. However, for both groups of children, only interlistener variability made a significant independent contribution to the model. The overall R^2 accounted for by the model was 18% for children in the SMI group and 29.6% for children in the TD group.

(d) How good is test–retest reliability for listeners of SMI and TD children? Descriptive data are shown in Figure 3 and suggest that intelligibility scores were very similar, though slightly higher for the second presentation than for the first within each of the two groups of children. Inferential statistics indicated that the difference between intelligibility scores on first and second presentations was significant for the SMI group (mean difference = 3.32, 95% CI [2.28, 4.36]), $t(119) = 6.33$, $p < .001$, $d = .57$; and for the TD group (mean difference = 1.68, 95% CI [0.76, 2.26]), $t(99) = 3.64$, $p < .001$, $d = .36$. Figure 4 shows a scatterplot of intelligibility scores by group for the first versus the second presentation of stimuli. Cohen's kappa values comparing line-by-line consistency within listeners for the first and second presentations of stimulus utterances were .881 ($p < .001$) for the SMI group and .851 ($p < .001$) for the TD group.

Figure 2. Intra- and interlistener variability by intelligibility score for children in typically developing (TD) and speech motor impairment (SMI) groups.



Question 2: Differentiating Children Who Have SMI From TD Children on the Basis of Intelligibility Scores

Results of the four sets of analyses are presented by specific question. Figure 5 shows intelligibility distributions for children in the SMI and TD groups.

(a) How accurate are intelligibility scores globally in differentiating between and classifying TD and SMI children? Results indicated that the area under the ROC curve was .98 ($p < .001$), 95% CI [.96, .99], indicating that intelligibility scores are excellent at separating children who have SMI from children who are TD (see Figure 6). From the coordinates of the ROC curve, we calculated sensitivity,

Table 2. Simultaneous multiple linear regression model against intelligibility for children with speech motor impairment (SMI) and those typically developing (TD).

Group	Predictor variable	Unstandardized *b* coefficient	Standardized β coefficient	*t*	*p* value	R^2 for model
SMI	Intralistener variability	−0.523	−.120	−1.379	.171	.180
	Interlistener variability	4.807	.442	5.062	.000	
TD	Intralistener variability	−0.051	−.033	−0.382	.703	.296
	Interlistener variability	−1.718	−.540	−6.322	.000	

Figure 3. Mean intelligibility scores for first and second presentations of a subset of intelligibility materials by group. SMI = speech motor impairment; TD = typically developing.

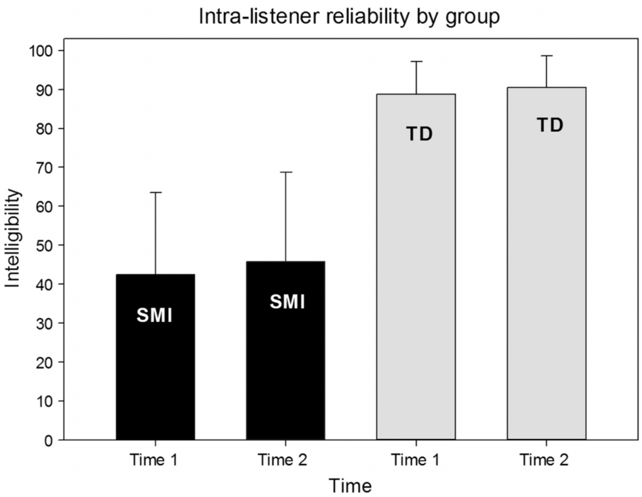
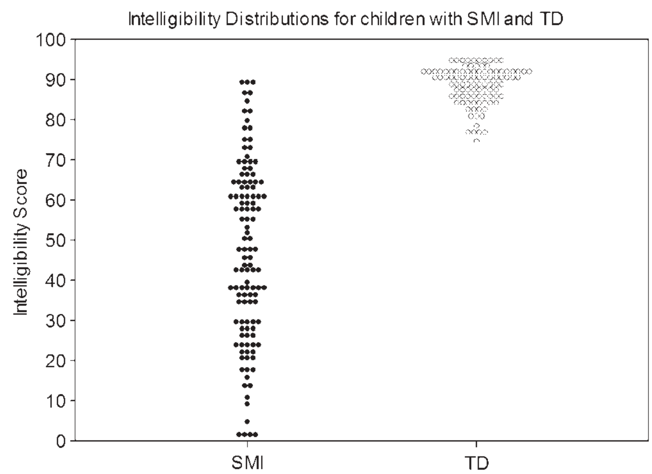


Figure 5. Distributions of intelligibility scores for children in the typically developing (TD) and speech motor impairment (SMI) groups.



specificity, and positive and negative likelihood ratios for intelligibility scores at 5% intervals (see Table 3).

Positive likelihood ratios provide an index of how likely a child is to have SMI at a given intelligibility score and are useful for ruling in the presence of a condition. Larger positive likelihood ratios indicate greater probability of a condition. Positive likelihood ratios above 10.0 are considered good indicators that a person has a condition (Dollaghan, 2007; Dollaghan & Horner, 2011). Negative likelihood ratios provide an index of how likely a child is to be TD at a given intelligibility score and are useful for ruling out the presence of a condition (in this case, SMI). Smaller negative likelihood ratios indicate lesser probabilities of a condition. Negative likelihood ratios below .10 are considered good indicators that a person does not have a condition (Dollaghan, 2007). Data shown in Table 3 indicated that positive likelihood ratios above 10.0 were

Figure 4. Intelligibility scores by group and first and second presentations of intelligiblity materials. SMI = speech motor impariment; TD = typically developing.

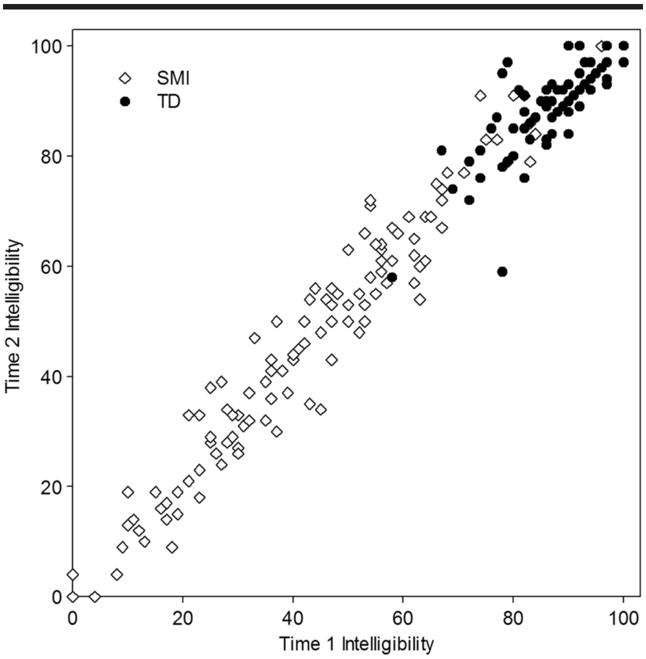


Figure 6. Receiver operating characteristic (ROC) curve for speech intelligibility scores for all children.

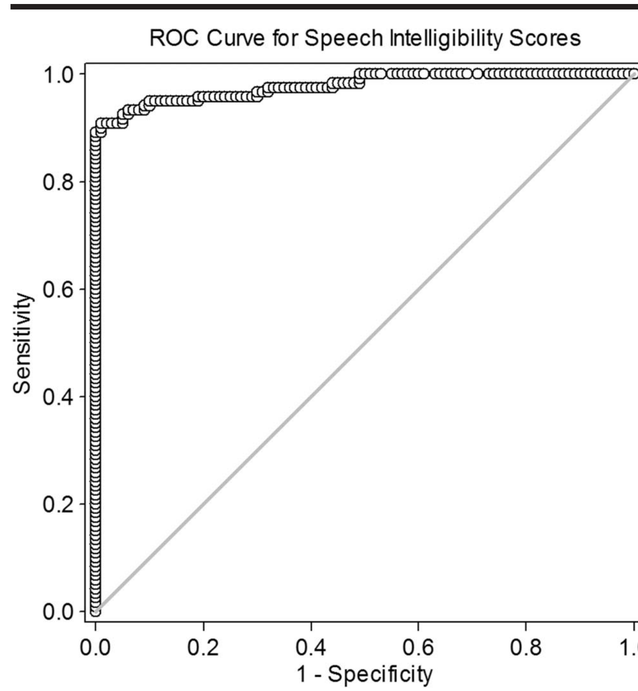


Table 3. Diagnostic accuracy of intelligibility scores on the basis of receiver operating characteristic (ROC) analyses.

Intelligibility score	Sensitivity	Specificity	LR+[a]	LR–[b]
10	0.05	1.00	—	0.95
15	0.08	1.00	—	0.93
20	0.12	1.00	—	0.88
25	0.20	1.00	—	0.8
30	0.28	1.00	—	0.73
35	0.43	1.00	—	0.69
40	0.48	1.00	—	0.58
45	0.54	1.00	—	0.52
50	0.54	1.00	—	0.46
55	0.58	1.00	—	0.43
60	0.66	1.00	—	0.34
65	0.78	1.00	—	0.22
70	0.86	1.00	—	0.14
75[c]	0.89	0.99	89.17	0.11
76	0.91	0.99	90.83	0.09
77	0.91	0.96	22.71	0.10
78	0.93	0.95	18.50	0.08
79	0.93	0.94	15.42	0.08
80	0.93	0.94	15.56	0.07
81	0.93	0.92	11.67	0.07
82	0.94	0.91	10.46	0.06
83	0.95	0.87	7.31	0.06
84	0.95	0.83	5.59	0.06
85	0.96	0.79	4.56	0.05
86	0.96	0.73	3.55	0.06
87	0.98	0.68	3.05	0.04
88	0.98	0.61	2.50	0.04
89	0.98	0.56	2.23	0.03
90[d]	0.99	0.51	2.02	0.02
95[e]	1.00	0.08	1.09	0.00

Note. For cells containing —, likelihood ratios could not be computed because the denominator was zero.

[a]LR+ refers to positive likelihood ratio. [b]LR– refers to negative likelihood ratio. [c]Reflects the lowest intelligibility score for typically developing children. [d]Reflects highest intelligibility score for children with speech motor impairment. [e]Reflects highest intelligibility score for typically developing children.

observed for intelligibility scores below 82%, indicating that children with intelligibility scores in this range had a very high likelihood of being SMI. Also, negative likelihood ratios below 0.10 were observed for intelligibility scores of 77% and above, indicating a low likelihood of being SMI. However, the range of intelligibility score between 76% and 82% yielded conflicting results, with concurrent occurrence of both high positive likelihood ratios (between 10.46 and 90.83), suggesting high probabilities of SMI and at the same time low negative likelihood ratios (between 0.09 and 0.06), suggesting low probabilities of SMI.

(b) Are there specific cutpoints in which intelligibility scores perform better at differentiating between groups of children? Results of binary logistic regression showed that intelligibility was a significant predictor of group membership (Wald $\chi^2 = 36.685$; $p < .001$; Nagelkerke $R^2 = .870$). Overall classification accuracy was 93.6%. Table 4 shows classification accuracy by intelligibility range. Results indicate that classification into groups was 100% accurate for intelligibility scores below 71% and over 89%.

Discussion

This study examined variability of speech intelligibility scores and how well intelligibility scores predicted dysarthria status among 5-year-old children with dysarthria secondary to CP (SMI; $n = 24$) and a control group of TD children ($n = 20$). Children produced a standard corpus of words and sentences ranging from one to four words in length. Five different listeners made orthographic transcriptions of each child, and no listener heard more than one child. From these data, we were able to examine variability within individual listeners across the different stimuli that they heard for the same child, as well as variability between different listeners who heard the same child. Several key results emerged from this study. First, variability associated with listeners makes a significant contribution to explaining the variance in intelligibility scores for children who are TD and children with SMI, although the magnitude of this contribution was not the same for the two groups of children. Second, listeners showed very small improvements in intelligibility when presented with the same stimuli more than once, but intralistener reliability was generally very consistent. As a final point, intelligibility scores appeared to differentiate very well between children who have SMI and those who are TD when intelligibility is at or below approximately 75% and above approximately 85%. Each of these findings is further discussed below.

Variability of Intelligibility Scores

We were interested in two different sources of variability related to listeners in this study. The first was intralistener variability, defined as performance differences within a given listener across different utterances produced by one child. Our measure of intralistener variability deliberately captured the impact of stimulus length within a given listener, which is known to be an important variable that affects intelligibility (Allison & Hustad, 2014; Hustad et al., 2012). The second was *interlistener variability*, defined as performance differences between all of the listeners ($n = 5$ in this study) who heard the same child. Our measure of interlistener variability also captured the impact of stimulus length across listeners to ensure that our measures of intra- and interlistener variability were conceptually and psychometrically comparable.

Our findings showed that intralistener variability was significantly greater than interlistener variability for listeners of children with SMI (about 4%) and for listeners of TD children (about 3%). One explanation for this finding relates to the stimulus material that children produced. Specific linguistic characteristics of speech stimuli can have a significant impact on listener performance; for example, words tend to be less intelligible than sentences (Hustad et al., 2012; Miller, Heise, & Lichten, 1951; Sitler, Schiavetti, & Metz, 1983). One reason for this phenomenon may be that listeners are able to apply their linguistic knowledge, including semantic and syntactic expectations, more readily to a sentence transcription task than to a word transcription task.

Table 4. Classification accuracy of intelligibility scores on the basis of logistic regression analyses.

Intelligibility range	Number of observations	Correct classification	Incorrect classification	Classification accuracy (%)
1–70	105.00	105.00	0.00	100.00
71–75	5.00	4.00	1.00	80.00
76	3.00	0.00	3.00	0.00
77	3.00	2.00	1.00	66.67
78	1.00	1.00	0.00	100.00
79	1.00	0.00	1.00	0.00
80	2.00	2.00	0.00	100.00
81	1.00	1.00	0.00	100.00
82	5.00	3.00	2.00	60.00
83	4.00	4.00	0.00	100.00
84	5.00	4.00	1.00	80.00
85	6.00	6.00	0.00	100.00
86	7.00	6.00	1.00	85.71
87	8.00	7.00	1.00	87.50
88	6.00	5.00	1.00	83.33
89	8.00	6.00	2.00	75.00
90	11.00	11.00	0.00	100.00
91–95	39.00	39.00	0.00	100.00

Our results suggest that stimulus characteristics induce variation in intelligibility measures within a given listener that has a considerable impact on performance but that this effect is generally fairly consistent among different listeners of a given child (hence, lower interlistener variability than intralistener variability). That is, the impact of stimulus characteristics is similar across different listeners for both TD children and children with SMI. However, when we examined the extent to which intra- and interlistener variability accounted for variance in intelligibility scores, we found that only interlistener variability made a significant contribution. This was true for both TD children and children with SMI. This finding suggests that differences among listeners may be more important than differences within individual listeners when it comes to intelligibility performance.

When we examined repeated performance on the same stimuli produced by the same child and transcribed by the same listeners, which were presented twice, randomly interspersed (intralistener reliability), we found that listeners showed small but significant improvements in intelligibility scores on the second presentation for both groups of children. Again, the magnitude of the difference was descriptively greater for children with SMI (about 3%) than for TD children (about 2%). This finding of improvement over repeated listening suggests that utterance characteristics may not be the only variable that contributes to intralistener variability. Listeners seem to learn, even with no instruction or feedback, a finding consistent with studies of listener learning in adult dysarthria (Borrie, McAuliffe, & Liss, 2012; Hustad & Cahill, 2003; Kim & Nanney, 2014). Note that although listeners appear to learn, the magnitude of change is small and probably not clinically meaningful. For both groups of children, Cohen's kappa values, which quantified utterance-by-utterance consistency between the first and second presentation within each listener, were very high and were similar in magnitude for both groups of children. This finding provides evidence that intralistener reliability was very good for listeners of children with SMI and TD children.

Our findings also showed that variability was greater for listeners of children with SMI than for listeners of TD children. This was true for both types of listener variability that we examined (intra- and inter). This finding is not surprising given that the speech signal of children with SMI contains more irregularities and thus may be less predictable in its acoustic characteristics (e.g., phonetic errors, temporal features, and spectral features) to listeners than typical speech patterns of 5-year-old children. However, although listeners of children with SMI had greater variability in their performance than listeners of TD children, results showed that interlistener variability accounted for more of the variance in intelligibility scores in TD children (29.5%) than in children with SMI (17%). This finding is difficult to explain but again may relate to irregularities in the speech signal of children with SMI and to the heterogeneity of production feature differences among children with SMI. That is, because production features are expected to be more heterogeneous among children with SMI than among TD children, it is very likely that there are more variables related to the speech signal that account for the variance in intelligibility scores. One result may be that listener variables, although clearly very important, play a smaller role in explaining intelligibility for children with SMI than for TD children, highlighting the importance of studies seeking to understand the impact of production characteristics on intelligibility of children with SMI.

Differentiating Children Who Have SMI From TD Children on the Basis of Intelligibility Scores

To determine how well intelligibility scores differentiated between children with SMI and TD children, we examined overall test accuracy via ROC analysis. We then completed a logistic regression to determine how well intelligibility scores predicted group membership. From these two analyses, we were able to examine likelihood ratios and classification accuracy for specific intelligibility benchmarks in determining whether a child had SMI or was TD.

Our results were very promising, suggesting that classification accuracy across the full range of intelligibility scores was approximately 95%. However, classification accuracy was 100% for children with intelligibility scores at or above 90 and for children with intelligibility scores at or below 70. The region from 70% to 90% intelligibility had some variability with regard to classification accuracy. To refine our understanding of these data, we examined likelihood ratios. Data shown in Table 3 indicated that positive likelihood ratios above 10.0 were observed for intelligibility scores below 82%; negative likelihood ratios below .10 were observed for intelligibility scores of 76% and above. However, intelligibility scores between 76% and 82% had concurrent occurrence of both high positive likelihood ratios (between 10.46 and 90.83), suggesting high probabilities of SMI and at the same time low negative likelihood ratios (between .09 and .06), suggesting low probabilities of SMI. The finding of a zone of uncertainty between 76% and 82% intelligibility was further corroborated by examination of classification accuracy by intelligibility benchmark (see Table 4). Results indicate that classification was least accurate (56%) for intelligibility scores from 76% to 82%. Together, our findings suggest that intelligibility scores appear to be a strong measure for differentiating children who are TD from children with SMI. This is especially true for children with intelligibility scores below 75% and above 85%.

Limitations and Future Directions

There are several important limitations to the present study. First, a relatively small number of children contributed speech samples for the study. This is particularly important in the context of the TD children (*n* = 20) who form the basis for the range of typical expectations for 5-year-old children. Within this sample, typical children ranged in their intelligibility between 75% and 95% (*M* = 88.6%; *SD* = 4.75%). Examining a larger number of children would help to establish more firmly a set of normative expectations.

Children in this study were an average age of 60.50 months (*SD* = 5.97 months). Additional sampling points spanning a wide range of ages are necessary to advance our understanding of age-specific expectations for intelligibility development.

We used a recitation format intelligibility task, the TOCS+ (Hodge & Daniels, 2007), in which listeners only had access to auditory information as they orthographically transcribed speech samples. Although the TOCS+ is a clinical tool with established validity (M. M. Hodge & Gotzke, 2014a; M. Hodge & Gotzke, 2014b), there are other ways that intelligibility can be measured. Previous work suggests that parent estimates of intelligibility are closely related to results obtained from orthographic transcription (Hustad et al., 2012). It may be worthwhile to examine variability among these types of estimates to determine the validity of such a measurement approach for identifying children who may fall outside of the range of typical expectations for intelligibility development.

Results of the present study showed that at the age of 5 years, children with intelligibility below 75% are extremely unlikely to be TD. Also, children with intelligibility over 85% are extremely likely to be TD. There is a gray area between 75% and 85% intelligibility where it is difficult to determine whether a child is typical or not on the basis of intelligibility scores alone. Once again, further study with a larger number of children spanning a wide range of ages is necessary to validate this finding and to determine the extent to which similar ambiguous ranges may exist at other ages.

Intelligibility deficits occur in many pediatric populations. Studies examining children with other speech disorders are necessary to determine whether findings generalize to other populations of children.

Clinical Implications

This study revealed several clinically relevant findings. First, intelligibility seems to be a useful tool for differentiating between TD children and children with SMI at 5 years of age. This measure is simple, practical, and ecologically valid in terms of its clinical application. Using a recitation format with utterances ranging from one to four words in length, we obtained results of the present study that suggest a tentative range of typical expectation for intelligibility of 5-year-old children is 75%–95% (*M* = 88.6; *SD* = 4.75). Note, however, that there was considerable variability within and between listeners who heard each child and that between-listener variability accounted for significant variance in intelligibility scores. Although determining how many listeners are necessary to capture the range of variability in children was beyond the scope of the present article, our results support the idea that more than one listener is necessary to ensure the validity of an intelligibility measure.

In the present study, children who had intelligibility below 75% were extremely likely to have a clinical diagnosis of dysarthria, while children with intelligibility above 85% were extremely likely to be TD. However, the range between 75% and 85% was an uncertain area that included a small number of children who were typical and a small number of children with dysarthria. Thus, one guideline might be to obtain a comprehensive speech-language evaluation for children with intelligibility scores below 85% at 5 years of age. This would ensure that children whose speech intelligibility is questionable or on the edge receive treatment as appropriate to ensure that they have the opportunity to maximize their functional communication ability.

Acknowledgments

This study was funded by Grant R01DC009411, awarded to the first author, from the National Institute on Deafness and Other Communication Disorders. Support was also provided by the

Waisman Center Core Grant P30HD03352 from the National Institute of Child Health and Human Development. The authors thank the children and their families who participated in this research and the graduate and undergraduate students at the University of Wisconsin–Madison who assisted with data collection and data reduction.

References

Allison, K. M., & Hustad, K. C. (2014). Impact of sentence length and phonetic complexity on intelligibility of 5-year-old children with cerebral palsy. *International Journal of Speech-Language Pathology, 16,* 396–407. doi:10.3109/17549507.2013.876667

Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech, Language, and Hearing Research, 55,* 290–305. doi:10.1044/1092-4388(2011/10-0349)

Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2012). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes, 27,* 1039–1055. doi:10.1080/01690965.2011.610596

Borrie, S. A., McAuliffe, M. J., Liss, J. M., O'Beirne, G. A., & Anderson, T. J. (2013). The role of linguistic and indexical information in improved recognition of dysarthric speech. *The Journal of the Acoustical Society of America, 133,* 474–482. doi:10.1121/1.4770239

Carrow-Woolfolk, E. (1999). *Test for Auditory Comprehension of Language* (3rd ed.). Austin, TX: Pro-Ed.

Choe, Y. K., Liss, J. M., Azuma, T., & Mathy, P. (2012). Evidence of cue use and performance differences in deciphering dysarthric speech. *The Journal of the Acoustical Society of America, 131,* EL112–E118. doi:10.1121/1.3674990

Coplan, J., & Gleason, J. R. (1988). Unclear speech: Recognition and significance of unintelligible speech in preschool children. *Pediatrics, 82,* 447–452.

Dickinson, H., Parkinson, K., Ravens-Sieberer, G., Thyen, U., Arnaud, C., Beckung, E., . . . Colver, A. (2007). Self-reported quality of life of 8–12-year-old children with cerebral palsy: A cross-sectional European study. *The Lancet, 369,* 2171–2178.

D'Innocenzo, J., Tjaden, K., & Greenman, G. (2006). Intelligibility in dysarthria: Effects of listener familiarity and speaking condition. *Clinical Linguistics & Phonetics, 20,* 659–675.

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore, MD: Brookes.

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research, 54,* 1077–1088. doi:10.1044/1092-4388(2010/10-0093)

Flipsen, P. (2006). Measuring the intelligibility of conversational speech in children. *Clinical Linguistics & Phonetics, 20,* 303–312.

Fudala, J. B. (2001). *Arizona Articulatory Proficiency Scale* (3rd ed.). Los Angeles, CA: Western Psychological Services.

Hodge, M., & Daniels, J. (2007). *TOCS+ Intelligibility Measures*. Edmonton, Canada: University of Alberta.

Hodge, M. M., & Gotzke, C. L. (2014a). Construct-related validity of the TOCS measures: Comparison of intelligibility and speaking rate scores in children with and without speech disorders. *Journal of Communication Disorders, 51,* 51–63. doi:10.1016/j.jcomdis.2014.06.007

Hodge, M., & Gotzke, C. L. (2014b). 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,* 417–426. doi:10.3109/17549507.2014.930174

Hustad, K. C., Allison, K. M., McFadd, E. D., & Riehle, K. (2013). Speech and language development in 2-year-old children with cerebral palsy. *Developmental Neurorehabilitation, 17,* 167–175.

Hustad, K. C., & Cahill, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology, 12,* 198–208.

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(2010/09-0176)

Hustad, K. C., Schueler, B., Schulz, L., & DuHadway, C. (2012). Intelligibility of 4-year-old children with and without cerebral palsy. *Journal of Speech, Language, and Hearing Research, 55,* 1177–1189.

Hustad, K. C., & Weismer, G. (2007). A continuum of interventions for individuals with dysarthria: Compensatory and rehabilitative approaches. In G. Weismer (Ed.), *Motor speech disorders* (pp. 261–303). San Diego, CA: Plural.

Kent, R., Weismer, G., Kent, J., & Rosenbek, J. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54,* 482–499.

Kim, H., Hasegawa-Johnson, M., & Perlman, A. (2011). Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica Logopaedica, 63,* 187–194. doi:10.1159/000318881

Kim, H., & Nanney, S. (2014). Familiarization effects on word intelligibility in dysarthric speech. *Folia Phoniatrica Logopaedica, 66,* 258–264. doi:10.1159/000369799

Lam, J., & Tjaden, K. (2013). Intelligibility of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research, 56,* 1429–1440. doi:10.1044/1092-4388(2013/12-0335)

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,* 408–416. doi:10.3109/17549507.2014.927922

Lindblom, B. (1990). On the communication process: Speaker-listener interaction and the development of speech. *Augmentative and Alternative Communication, 6,* 220–230.

Liss, J. M., Spitzer, S., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 112,* 3022–3030.

McAuliffe, M. J., Carpenter, S., & Moran, C. (2010). Speech intelligibility and perceptions of communication effectiveness by speakers with dysarthria following traumatic brain injury and their communication partners. *Brain Injury, 24,* 1408–1415. doi:10.3109/02699052.2010.511590

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41,* 329–335.

Montag, J. L., AuBuchon, A. M., Pisoni, D. B., & Kronenberger, W. G. (2014). Speech intelligibility in deaf children after long-term cochlear implant use. *Journal of Speech, Language, and Hearing Research, 57,* 2332–2343. doi:10.1044/2014_JSLHR-H-14-0190

Palisano, R., Rosenbaum, P., Walter, S., Russell, D., Wood, E., & Galuppi, B. (1997). Development of the Gross Motor Function Classification System. *Developmental Medicine & Child Neurology, 39,* 214–223.

Paneth, N., Hong, T., & Korzeniewski, S. (2006). The descriptive epidemiology of cerebral palsy. *Clinics in Perinatology, 33,* 251–267.

Pennington, L., Roelant, E., Thompson, V., Robson, S., Steen, N., & Miller, N. (2013). Intensive dysarthria therapy for younger children with cerebral palsy. *Developmental Medicine & Child Neurology, 55,* 464–471. doi:10.1111/dmcn.12098

Rong, P., Loucks, T., Kim, H., & Hasegawa-Johnson, M. (2012). Relationship between kinematics, F2 slope and speech intelligibility in dysarthria due to cerebral palsy. *Clinical Linguistics & Phonetics, 26,* 806–822. doi:10.3109/02699206.2012.706686

Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37,* 55–63.

Schneider, J. W., Gurucharri, L. M., Gutierrez, L. M., & Gaebler-Spira, D. J. (2001). Health-related quality of life and functional outcome measures for children with cerebral palsy. *Developmental Medicine & Child Neurology, 43,* 601–608.

Sitler, R. W., Schiavetti, N., & Metz, D. E. (1983). Contextual effects in the measurement of hearing-impaired speakers' intelligibility. *Journal of Speech and Hearing Research, 26,* 30–34.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55,* 779–798.

Turner, G. S., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 38,* 1001–1013.

Weismer, G. (2008). Speech intelligibility. In M. J. Ball, M. R. Perkins, N. Muller, & S. Howard (Eds.), *The handbook of clinical linguistics* (pp. 568–582). Oxford, United Kingdom: Blackwell.

Weismer, G., Jeng, J. Y., Laures, J. S., Kent, R. D., & Kent, J. F. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica et Logopaedica, 53,* 1–18.

Weismer, G., & Martin, R. (1992). Acoustic and perceptual approaches to the study of intelligibility. In R. Kent (Ed.), *Intelligibility in speech disorders* (pp. 67–118). Philadelphia, PA: John Benjamins.

Weiss, C. (1982). *Weiss Intelligibility Test*. Tigard, OR: CC Publications.

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

Yunusova, Y., Weismer, G., Kent, R. D., & Rusche, N. M. (2005). Breath-group intelligibility in dysarthria: Characteristics and underlying correlates. *Journal of Speech, Language, and Hearing Research, 48,* 1294–1310.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale* (4th ed.). San Antonio, TX: The Psychological Corporation.