Journal of Speech and Hearing Research, Volume 35, 1095–1104, October 1992

Intelligibility Assessment in Developmental Phonological Disorders: Accuracy of Caregiver Gloss

Joan Kwiatkowski
Lawrence D. Shriberg
Department of Communicative Disorders and The Waisman Center on Mental Retardation and Human Development
University of Wisconsin, Madison

Fifteen caregivers each glossed a simultaneously videotaped and audiotaped sample of their child with speech delay engaged in conversation with a clinician. One of the authors generated a reference gloss for each sample, aided by (a) prior knowledge of the child's speech-language status and error patterns, (b) glosses from the child's clinician and the child's caregiver, (c) unlimited replays of the taped sample, and (d) the information gained from completing a narrow phonetic transcription of the sample. Caregivers glossed an average of 78% of the utterances and 81% of the words. A comparison of their glosses to the reference glosses suggested that they accurately understood an average of 58% of the utterances and 73% of the words. Discussion considers the implications of such findings for methodological and theoretical issues underlying children's moment-to-moment intelligibility breakdowns during speech-language processing.

KEY WORDS: phonology, intelligibility, assessment, caregivers, glossing

There currently is no comprehensive theoretical perspective to explain why certain children with speech delays are more difficult to understand than others and why the same child is intelligible during one utterance and quite unintelligible on the next. Findings from several studies indicate that the severity of a child's speech involvement, as indexed by a Percentage of Consonants Correct (PCC) score in continuous speech, accounts for only 20% of the variance in children's intelligibility (cf. Bishop & Edmundson, 1987; Shriberg, Kwiatkowski, Best, Hengst, & Terselic-Weber, 1986). The remaining variance has been associated with a child's specific pattern of error types, productive language status, and profile of prosody-voice involvement (Shriberg & Kwiatkowski, 1982). Weston and Shriberg (in press) report that structural and linguistic variables, including certain syllable shapes, grammatical classes, and stress conditions, are significantly more likely to be associated with unintelligible speech when compared to their usual frequency of occurrence. All such findings may be associated with both speaker and listener variables. Moreover, they do not explain the utterance-to-utterance variability in intelligibility that is typical of children with moderately delayed to severely delayed speech acquisition.

A major problem in accounting for variable intelligibility in this clinical population is the difficulty of intelligibility assessment. Of foremost concern is the validity and the clinical-research utility of word-level or sentence-level protocols compared to analysis procedures based on a sample of conversational speech.

Word-Level and Sentence-Level Intelligibility Assessment

Word-level and sentence-level intelligibility measures have been developed for several communication disorders in children and adults, including deafness (Monsen,

© 1992, American Speech-Language-Hearing Association

0022-4685/92/3505-1095$01.00/0

1981), dysarthria (Kent, Weismer, Kent, & Rosenbek, 1989; Yorkston & Beukelman, 1981), and articulation disorders (Weiss, 1982: Part A). Each approach uses citation form speech, including spontaneous production, imitation, or reading, with speakers' responses judged by either one or a panel of listeners whose glosses determine the intelligibility of each response. As described by Kent et al. (1989, p. 495), advantages to single word intelligibility testing include "(a) quantification in terms of percentage correct, (b) [the potential for] a phonetic feature analysis of the errors, (c) significant elimination of syntactic, prosodic, and other variables that affect sentence production or conversation, and (d) a simple response from subjects." A primary question in intelligibility assessment for developmental phonological disorders is whether the variables listed under (c) in the above quotation are sources of error variance or whether, in fact, they play the central role in explanatory models. This question will be addressed later in the paper. Here, it is important to consider a number of threats to the validity of single-word or sentence-level intelligibility measures when applied to young children with delays in speech acquisition.

A major validity constraint is the significant difference in articulation reported in many studies comparing citation form speech to conversational speech (cf. review of literature in Morrison & Shriberg, 1992). At the phonemic level, findings from a recent study of 61 children with speech delays indicate that earlier-developing sounds (nasals, glides) are produced more accurately in conversational speech, whereas later-developing sounds (fricatives, liquids) are produced more accurately in citation forms (Morrison & Shriberg, 1992). Additional analyses of citation forms indicate that children often make unnatural or exaggerated articulatory movements on both consonants and vowels—behaviors that confound the interpretation of both acoustic and perceptual analyses of the articulatory loci of unintelligible speech. Such gestures are especially apparent when responses are obtained by imitation, particularly when children in an intervention program have a generalized response to examiners characterized by the attempt to use "good speech." Klein (1984), Dyson and Robinson (1987), and Morrison and Shriberg report that compared to conversational speech, word-level tasks may also yield significantly lower percentages of initial and final consonant deletions, neutralizations, stopping, and weak syllable deletions. Hodson and Paden (1981) report that such error patterns (i.e., stridency deletion, stopping, final consonant deletion, and syllable reduction) tend to occur frequently in unintelligible children with speech delays compared to error patterns found in intelligible children with speech delays. Thus, there appears to be considerable empirical evidence that for the purpose of intelligibility generalizations for children with speech delays, articulation in citation forms differ significantly from articulation during conversational speech.

A second set of factors constraining the concurrent validity of citation form tests of intelligibility concern the limitations imposed by the cognitive and language levels of these children. Approximately 80% of children with developmental phonological disorders of unknown origin have delays in productive language; 10% to 40% also have delays in language comprehension (Shriberg & Kwiatkowski, 1988). Ideally, as proposed by Kent et al. (1989), minimal word pairs selected for intelligibility assessment should include acoustic-phonemic contrasts that are important for correct glossing by listeners. Development of a suitable set of such stimuli for children with speech delays is constrained by the paucity of minimally contrastive words within these children's comprehension and production vocabularies. Nonsense words would not be appropriate; these children's articulation of nonsense words result in types and distributions of speech sound errors that are not typical of performance on known words (cf. Shriberg & Kwiatkowski, 1980). Moreover, evidence from a number of sources suggests that structural and grammatical characteristics of words and utterances are associated with different articulation outcomes in both children acquiring speech normally and children with speech delays. For example, Klein and Spector (1985) found that children with speech delays produce atypical speech errors in the syllables of polysyllabic words that contain reduced stress, and they include syllables in polysyllabic words in imitation that they do not include when producing words spontaneously. Camarata and Leonard (1986) reported that very young normally developing children produce object words more accurately than action words; accordingly, a preponderance of nouns in an intelligibility protocol might overestimate the intelligibility of young children with delayed speech (cf. Campbell & Shriberg, 1982; Shriberg & Kwiatkowski, 1980). Panagos and colleagues have reported increases in segmental errors as the length and syntactic complexity of a sentence increases (Panagos & Prelock, 1982; Panagos, Quine, & Klich, 1979; Prelock & Panagos, 1989, 1991). Thus, there are major constraints on the structural, canonical, semantic, and grammatical characteristics and distributions of words and sentences that might be candidates for word- or sentence-level intelligibility protocols.

Intelligibility Assessment Based on Conversational Speech

Methodological problems associated with intelligibility assessment using continuous speech samples (e.g., Shriberg & Kwiatkowski, 1982; Weiss, 1982: Part B) can be divided into three concerns. First, by definition, spontaneous speech sampling includes the possibility that variance in intelligibility estimates could be biased by the particular quantitative and qualitative linguistic content spontaneously produced in a sample. In a series of studies of conversational speech sampling (Morrison & Shriberg, 1992; Shriberg, 1986; Shriberg & Kwiatkowski, 1980, 1982, 1985), conversational speech samples have yielded stable and representative distributions of grammatical content, canonical forms, and intended phonemes. Moreover, clinical and research experience with more than 400 3- to 6-year-old children with developmental phonological disorders of unknown origin indicates that fewer than 10% cannot be engaged in spontaneous conversational speech or have speech that is too unintelligible for glossing by well-trained examiners (Shriberg, in press).

A second concern about conversational speech samples is raised when scaling procedures are used to estimate de-

grees of intelligibility or unintelligibility. Procedures that include equal-interval appearing scales have been critiqued on a number of psychophysical scaling and statistical criteria (Bernthal & Bankson, 1988; Schiavetti, Metz, & Sitler, 1981). Whether based on equal-interval appearing scales or direct magnitude estimation, intelligibility scaling typically requires averaged data from a panel of listeners, providing an estimate of overall intelligibility rather than a percentage of understood message units. However, even when researchers in child phonology have assessed intelligibility as a percentage of intelligible words, they have not demonstrated that the judges' glosses validly match the child's intended utterances.

The major problem in using conversational speech samples to examine moments of unintelligibility is that the intended targets of some unintelligible strings cannot be ascertained. Adult speakers generally can be asked to repeat exactly what was intended or even to write out their intended utterances. With children, requests to repeat utterances are typically not productive and are incompatible with the goals of free speech sampling. For these reasons, caregivers and others familiar with unintelligible speakers have routinely been enlisted in clinical and research environments to provide glosses for unintelligible strings. Goehl and Martin (1987) have provided limited data on the validity of such information. In a preliminary study of three mother-child pairs using a sentence repetition task, they found that caregivers were as much as 30% better than experienced speech clinicians at recognizing words spoken by their child. Generalizations from these interesting findings are obviously limited by the number of mother-child pairs and the use of imitation as the response mode.

A Rationale for Using Conversational Speech Samples

The primary consideration underlying the choice of alternative approaches to intelligibility assessment is the purpose for testing—to *classify* a speaker's level of unintelligibility, to *describe* certain linguistic dimensions of unintelligibility, or to attempt to *explain* the relevant correlates of unintelligibility. The authors take the position that for each of these goals, conversational speech sampling provides the only valid approach for children with phonological disorders of known and unknown origin: Whether obtained spontaneously or by imitation (reading cannot be used with preschool children), word- or sentence-level procedures are not valid for any of the three purposes of intelligibility assessment. If used to classify level of severity, they underestimate involvement because children delete more segments in the complex contexts of conversational speech (Morrison & Shriberg, 1992). If used to describe the linguistic dimensions of unintelligibility, word- and sentence-level protocols confound children's language-level difficulties with their speech-level constraints. Finally, if used to attempt explanatory models, word- or sentence-level protocols do not evoke the interaction of the language-speech-voice-prosody dimensions that underlie the moment-to-moment variability in intelligibility that typifies the conversational speech of children with developmental speech-language disorders.

The current study addresses the feasibility of using caregiver glosses of children with speech delays to aid clinical-research assessments of intelligibility in conversational speech. Three questions are posed: (a) What percentage of utterances and words do caregivers think they understand? (b) What percentage of utterances and words do caregivers accurately understand? (c) What strategies do caregivers use to gloss words that are difficult to understand?

Method

Children, Clinicians, and Caregivers

Conversational speech samples were obtained from 15 children with speech delays who were receiving services in a university-affiliated phonology clinic for management of speech delays of unknown origin. As shown in Table 1 the nine boys and six girls ranged in age from 3 years, 2 months to 5 years, 11 months. Severity of speech involvement ranged from moderate to severe, as indexed by their Percentage of Consonants Correct (Shriberg & Kwiatkowski, 1982; Shriberg, Kwiatkowski, Best, Hengst, & Terselic-Weber, 1986). Prosody-voice status based on the same continuous speech samples ranged from 0–100% appropriate utterances on each of six suprasegmentals (Phrasing, Rate, Stress, Loudness, Pitch, and Quality) scored on the Prosody-Voice Screening Profile (Shriberg, Kwiatkowski, & Rasmussen, 1990). Language production status, as indicated by Mean Length of Utterance and Structural Stage (Miller, 1981), ranged from within normal limits for chronological age to a delay of greater than one year.

The conversational speech samples for the 15 children were obtained by each child's graduate-level student clinician. Each of the 11 participating clinicians, including both first-year and second-year students, provided intervention services for one or two children. Conversational speech sampling was a part of the child's clinic routine for the purpose of monitoring generalization of target sounds to spontaneous speech. Thus, during the period when data were obtained for the study, children were familiar with the clinician and clinic routines, and the clinicians were familiar with the children's error patterns.

Fourteen of the caregivers were the children's mothers. One of the caregivers was a grandmother who was the child's primary caregiver during the more than 40 hours per week that the mother worked. The 15 caregivers represented a typical cross section of those seen in a university-affiliated speech-language clinic in the city of Madison, Wisconsin. No additional measures of family size or socioeconomic status were obtained.

Procedures

Data collection and data reduction proceeded in five stages: (a) obtaining a speech sample, (b) obtaining the caregiver gloss, (c) obtaining the clinician gloss, (d) deriving the reference gloss, and (e) calculating agreements and

TABLE 1. Demographic and speech-language production data for 15 children with speech delays.

			Speech-Prosody							Language	
				Prosody[a]			Voice[a]				
Child	Gender	Age yrs.: mos.	PCC[b]	Phrasing	Rate	Stress	Loudness	Pitch	Quality	*MLU*[c]	Structural[d] stage (mos.)
1	F	3:2	71.9	90	100	90	95	100	100	3.85	WNL
2	F	3:7	71.0	95	65	100	85	100	75	4.84	WNL
3	F	3:9	72.0	100	100	100	50	100	45	3.57	>6
4	F	4:1	59.5	100	100	100	90	100	100	3.54	>6
5	F	4:2	66.4	100	100	100	100	100	100	4.30	>6
6	F	5:0	52.4	85	84	100	100	100	95	4.84	>6
7	M	3:8	54.4	94	100	89	78	100	94	2.13	>12
8	M	3:9	73.4	85	90	100	65	100	80	4.24	WNL
9	M	3:10	78.9	85	80	40	100	100	95	5.23	WNL
10	M	4:1	73.2	80	95	100	75	100	0	3.89	>6
11	M	4:5	73.4	80	100	100	100	100	90	4.87	WNL
12	M	4:7	82.8	95	90	100	100	100	100	4.18	>6
13	M	4:7	74.2	80	95	100	100	100	100	5.31	>6
14	M	4:9	49.7	95	13	50	70	100	75	2.06	>12
15	M	5:11	44.9	100	95	20	100	100	100	2.40	>12

[a]Prosody-Voice Screening Profile (Shriberg, Kwiatkowski, & Rasmussen, 1990); entries for each suprasegmental are percentages of utterances scored as appropriate. [b]Percentage of Consonants Correct (Shriberg & Kwiatkowski, 1982); entries are percentages. [c]Mean Length of Utterance (Miller, 1981). [d]Miller (1981); WNL = Within Normal Limits; other entries are months delayed.

disagreements between the reference and the caregiver gloss.

The speech sample. Each child's clinician obtained a continuous speech sample using individualized materials and topics to evoke conversational speech (Shriberg & Kwiatkowski, 1985) and procedural and technical conventions for simultaneous videocassette and audiocassette taping (Shriberg, Kwiatkowski, & Snyder, 1989). During the taping, the clinician indicated her interpretation of the child's intended utterances via verbal and gestural responses. At times, she also verbally glossed portions of the child's utterances to facilitate subsequent written glossing.

The caregiver's written gloss. Prior to speech sampling, the study had been described to the caregivers as an attempt to learn more about intelligibility in children with speech delays by using caregiver glosses to identify why certain words were unintelligible to the student clinicians. The caregivers were used to such glossing tasks, as they had had prior experience helping the student clinicians to understand their child's speech. It was assumed that the caregivers would not be influenced by the clinicians' interpretations of their children's utterances during speech sampling because they had often heard the clinicians misgloss their children's speech. Support for this assumption was provided by the observation that there were no occasions of a caregiver changing her gloss after hearing the clinician's gloss.

Each child's caregiver observed the recording session from behind a one-way mirror in an adjacent room. Caregivers were positioned to have a clear view of both their child and the materials used to engage the child in conversation. They viewed all conversational materials through the mirror positioned behind the child, and they obtained a head and shoulders front view of the child from a large color video monitor in the viewing room positioned approximately 4 feet away. In order for caregivers to make maximal use of immediate contextual cues, they were asked to write down as many of the child's utterances as they could manage to record.

Following the taping session, the first author—the student clinicians' supervisor—viewed the video tape of the speech sample with the caregiver, who generated a word-by-word orthographic gloss of each of the child's utterances. The supervisor paused the tape after each utterance so that the caregiver would not hear the clinician's response before completing a gloss of each preceding utterance. The caregiver used the glosses written during the taping to confirm each of her glosses from the video tape. Video-playback was used to provide immediate contextual cues to aid glossing. Utterances were replayed on caregiver request or whenever portions of a succeeding utterance were inadvertently deleted due to operation of the pause function. Syllables that the caregiver could not gloss after replays were marked with an asterisk, using conventions described in Shriberg (1986) to gloss strings of unintelligible syllables. Fillers, nonspeech noises or sound effects, exclamations, part-word repetitions, and the frequently occurring words "yes" and "no" and phrase "I don't know" were not glossed.

The clinician's written gloss. Following the taping session, the child's clinician listened to the audiotaped speech sample and generated an orthographic gloss of the child's utterances using procedures similar to those for obtaining the caregiver's gloss. However, the clinician was allowed to hear her own taped glosses whenever they occurred after child utterances and was encouraged to make maximal use of all contextual cues. Video-playback to provide contextual cues was not necessary because glossing was completed within a few hours of the original taping.

The reference gloss. Using both the caregiver's and the clinician's glosses as an aid to the audio-playback, the first author simultaneously glossed and narrowly transcribed all

the speech samples. Speech segments and syllable counts were used during transcription to judge the accuracy of the glosses provided by the clinician and the caregiver to generate a new gloss when neither was appropriate. Unlimited replays of the audiotapes were permitted, and all other technical procedures and linguistic conventions followed those described in Shriberg (1986). Average intrajudge agreement for glossing was 94% for a 10% sample of words randomly selected from each of the 15 speech samples.

The reference gloss was regarded as the accurate gloss of the child's speech because to generate it, the supervisor (a) had used her prior knowledge of the child's speech-language status and speech-language error patterns, (b) had both the clinician's and the caregiver's gloss as a reference, (c) had the opportunity to replay each difficult word or utterance many times, and (d) had used the phonetic transcription process to gloss what may otherwise have been unintelligible. Grammatical morphemes were included in the gloss only when the morpheme was transcribed as marked in some way. The children's glosses of grammatical morphemes were consistent with what was known about their grammatical development from the language analysis summarized in Table 1. As discussed earlier, however, even these procedures could not guarantee that the reference glosses, in fact, reflected exactly what each child intended to say.

Data reduction. Totals for each of the following were obtained for each caregiver's orthographic gloss: (a) the number of words glossed by the caregiver, (b) the percentage of caregiver's glossed words that were considered *accurate glosses* because they agreed with the reference gloss, and (c) the percentage of caregiver's glossed words that were considered *inaccurate glosses* because they did not agree with the reference gloss. Accurate glosses were defined as exact agreements between caregiver and reference glosses on the occurrence of both intelligible and unintelligible words. Inaccurate glosses were defined as words that were unintelligible (marked with an asterisk) in only one of the glosses, words that were glossed differently in each source, intelligible or unintelligible words that were missing from one of the glosses, and words that represented close glosses (i.e., they differed only in presence/absence of a bound grammatical morpheme; they were different forms of the same root word [e.g., "have" versus "has"]; or they contained the same words but in different order).

In addition to the word-level coding, utterance-level coding of the caregiver's orthographic gloss was accomplished in two ways. First, each caregiver's utterance gloss was coded as *fully intelligible* (to the caregiver) when all words in the utterance were glossed, *partially intelligible* when at least one word, but not all words were glossed, and *unintelligible* when none of the words in the utterance were glossed. Second, the degree of agreement between each caregiver's gloss and the corresponding reference gloss was coded. On the assumption that the two glosses would be identical for children with normally developing speech, a caregiver's utterance gloss was coded as *fully intelligible* when the glosses for all of the words in the utterance were identical, *partially intelligible* when the two glosses for some but not all the words in the utterance were identical, and *unintelligible* when there was no agreement between the two glosses.

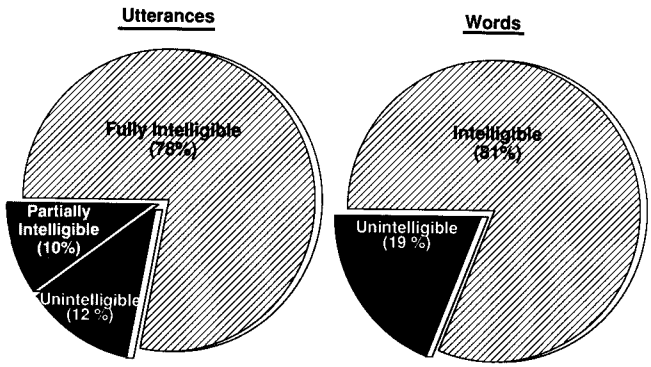

FIGURE 1. Mean percentages of utterances and words that caregivers thought were intelligible. The mean percentages of partially intelligible utterances and unintelligible utterances and words are also shown.

Thus, the two sets of word-level and utterance-level data reflect what caregivers thought they understood of their child's conversational speech and what they accurately understood. The data sets for the following questions were comprised of 3,530 words embedded in 877 utterances in the 15 transcripts. Due to the differences in language productivity across the 15 children, percentage scores were used in all analyses, with nonparametric statistics used for the inferential tests.

Results

Percentage of Words and Utterances Caregivers Thought They Understood

As shown by the cross-hatched sections in the circle graph in Figure 1, the two dependent variables, percentage of fully intelligible utterances (78%) and percentage of intelligible words (81%), yielded comparable findings, with caregivers on the average indicating they thought they understood approximately 80% of their child's conversational speech. Figure 2 includes these same data for each of the 15 caregiver-child pairs. The intelligibility status of children is arranged from left-to-right in descending order, based on utterance-level intelligibility. There is close correspondence between the utterance-level and the

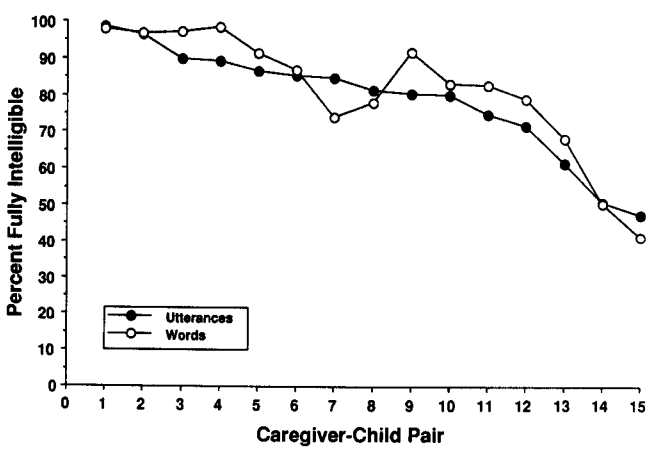

FIGURE 2. Percentages of fully intelligible utterances and words glossed by each of the 15 caregivers.

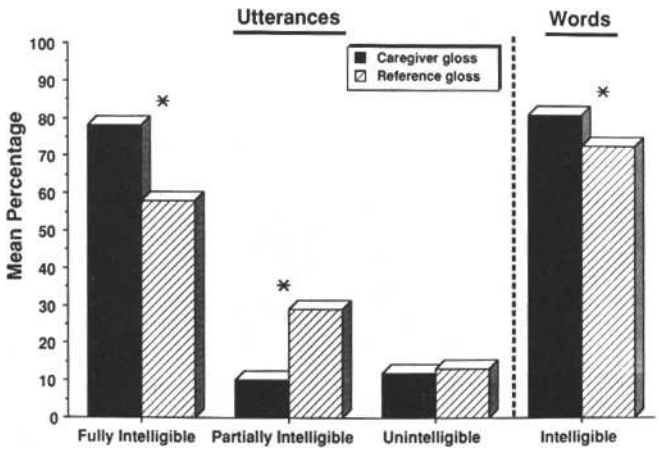

FIGURE 3. Mean percentages of utterances and words that caregivers thought they understood (caregiver gloss) and accurately understood (reference gloss). The percentages for partially intelligible and unintelligible utterances are also shown.

word-level accuracy data (i.e., reflecting intelligibility within and between utterances, respectively) across a wide range of intelligibility levels (42–98%). Thus, under conditions in which caregivers viewed both a live and recorded speech-sampling session and were allowed repeated replays of difficult-to-understand utterances, individual caregivers claimed to understand between nearly 100% to fewer than 50% of their child's utterances and words.

Percentage of Utterances and Words Caregivers Accurately Understood

Figure 3 includes comparison data for the accuracy of caregivers' glosses at the utterance level (left panel) and word level (right panel). As indicated by the asterisks in both panels, Wilcoxon Matched-Pair Signed-Ranks tests indicated that caregivers accurately understood significantly fewer utterances and words than they thought they understood. As reviewed previously, they thought they understood an average of 78% of utterances and 81% of words. Compared to the reference gloss, however, only an average of 58% (range: 29–95%) of these utterances were accurately glossed and an average of only 73% (range: 26–96%) of their intelligible words were accurately glossed. The findings for fully intelligible utterances affected the comparisons for partially intelligible utterances, with caregivers significantly underestimating the percentages for this latter category. Caregivers did not differ significantly from the respective reference glosses on the proportion of completely unintelligible utterances, averaging approximately 12–13% in the two sources.

Figure 4 contains the data for the individual caregiver-child pairs for the fully intelligible utterances, with percentages sorted left-to-right in descending order based on the caregiver gloss. Differences in the accuracy of the caregiver glosses for fully intelligible utterances ranged from 3 to 55% across the 15 transcripts. Accuracy was only generally

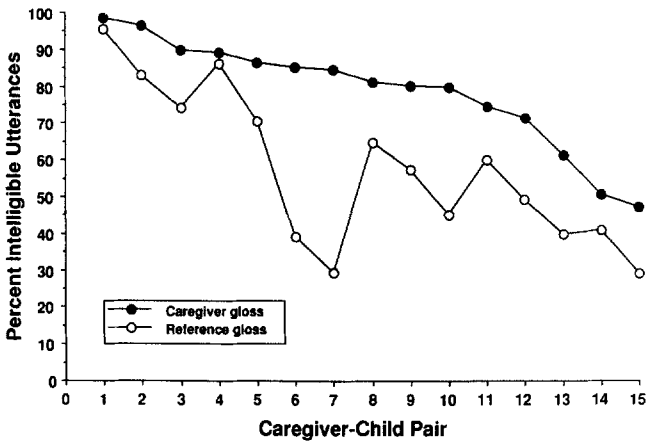

FIGURE 4. Percentages of fully intelligible utterances that caregivers thought they understood (caregiver gloss) and accurately understood (reference gloss).

related to level of intelligibility, as glossed by a child's caregiver. Specifically, the largest caregiver-reference differences occurred for pairs 6 and 7 and in general for a few other pairs in the middle of the distribution, with relatively more caregiver accuracy evident for the most and least intelligible children.

Differences in caregivers' ability to accurately gloss their child's utterances were not accounted for by the child's severity of speech involvement as indexed by Percentage of Consonants Correct (PCC). Spearman Rank-Order correlations for PCC with accurately glossed utterances and words were .20 and .43, respectively. Thus severity of speech involvement accounted for only 4% to approximately 19% of the variance in the accuracy of caregiver glosses.

Strategies Caregivers Use to Gloss Words That Are Difficult to Understand

Table 2 contains a percentage summary of the types of disagreement between the caregivers' glosses and the reference gloss. In this context, the term *accurately glossed* is used to identify words that were glossed exactly the same or were classified as unintelligible in both the caregiver and reference gloss. As shown at the top of the table, the 81% of

TABLE 2. Descriptive statistics for subcategories of words that were accurately glossed and not accurately glossed.

Word Category	*M*	*SD*	Range
Accurately Glossed			
Intelligible	71	21	25–96
Unintelligible	10	13	0–47
Not Accurately Glossed			
Unintelligible	5	6	0–20
Different	5	3	<1–13
Missed	4	4	0–17
Intelligible	2	2	0–10
Close	2	2	0–7
Added	1	1	0–4
Reversed	<1	1	0–4

Note. (*M*) Mean; (*SD*) Standard Deviation. All data are percentages.

words accurately glossed by caregivers included averages of 71% intelligible words and 10% unintelligible words. The wide dispersions from these averages for all the variables, as shown throughout the table in the standard deviation and range columns, demonstrate the individual differences in the data for caregiver-child pairs. On average, of the remaining 19% of words considered inaccurately glossed (a) 5% were unintelligible to the caregiver, but glossed in the reference; (b) 5% were glossed by caregivers as different words compared to the reference; (c) 4% were missed or deleted from the caregivers' gloss compared to the reference; (d) 2% were glossed by the caregiver, but considered unintelligible despite concerted efforts to gloss them in the reference; (e) 2% were considered close to the words in the reference gloss, approximately evenly divided between words that differed only on the presence/absence of a grammatical morpheme or were different forms of the same root word; (f) 1% were words the caregiver added to the gloss, but had no counterpart in the reference gloss; and (g) fewer than 1% were words that agreed with the reference gloss, but were sequenced differently. Thus, disagreements between the caregiver and reference glosses were spread among seven categories, with the ranges indicating that for some caregivers, disagreements included a fairly large percentage in at least one category.

Both the 5% of words that represented different glosses and the 2% of words that represented close glosses were suitable for identifying strategies caregivers might use to gloss words that are difficult to understand. The data sets consisted of 161 words (range: 3–29 words per caregiver) with different glosses and 79 words (range: 0–16 per caregiver) with close glosses. While all caregivers' glosses contained words that were glossed differently compared to the reference gloss, only 11 of the 15 caregivers' glosses contained words that were close to the reference gloss. Thus the limited number of available tokens is not appropriate for inferential statistical analysis, and the results of descriptive analyses are considered only suggestive.

A first prediction might be that when a word is initially unintelligible, the caregiver may use a word from the same word class as that intended by the child as the most likely gloss. In fact, as shown by the mean percentages and the standard deviation bars in the left panel in Figure 5, this was not the case. The left panel in Figure 5 includes the mean and standard deviations for the percentage of words categorized into a different word class than that by the child (i.e., as represented in the reference gloss) and into the same word class. When caregiver glosses disagreed, an average of 61% of their glosses were words from a different word class, with only 39% from the same word class. In the right panel are the percentages of words that caregivers categorized as beginning with a different sound versus the same sound as the child's intended or replacement sound as indicated in the transcription used to create the reference gloss. A second prediction might be that glosses that disagree will begin with the same initial sound as the sound produced by the child. As shown in the right panel in Figure 5, this was not observed. Rather, an average of 71% began with vowels or consonants that differed from the reference gloss, with only 29% beginning with the same initial sound. Thus, caregivers do not

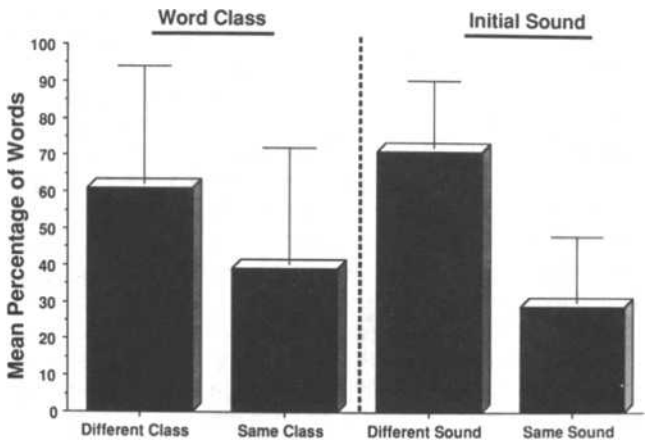

FIGURE 5. Mean percentages of caregivers' words (glossed differently from the reference gloss) categorized into those in a different or same word class and those beginning with a different or the same sound.

appear to be using either the same word class or same initial sound as a primary strategy to interpret difficult to understand words.

Close glosses were also analyzed for potential strategies to resolve why caregivers had not accurately glossed words for which they apparently had sufficient information to at least attempt a gloss. A strategy that appeared to be operating might be phrased as "credit the child with adult syntactic forms." Of the total of 79 words in the caregivers' glosses that were close to the reference gloss, 67 (85%) were accounted for by syntactic changes that included the (a) addition of appropriate grammatical morphemes (e.g., use of "here's" rather than "here"), (b) use of appropriate case forms for verbs (e.g., use of "has" rather than "have" for 3rd person singular), (c) use of appropriate negative forms (e.g., "doesn't" rather than "don't"), (d) inclusion of infinitives rather than catenatives (e.g., "going to" rather than "gonna"), and (e) use of more formal rather than casual word forms (e.g., "have" rather than "got"). In contrast, in only three glosses did caregivers replace more formal word forms with casual forms (e.g., "bicycle" versus "bike"). The remaining nine words could not be classified within any of the previous categories using available contextual cues. These words included different tense forms of the same verb (e.g., "know" versus "knew"), another form of the same modal (e.g., "can" versus "could"), a contracted form of the same words (e.g., "Where is" versus "Where's"), or a variant of the same lexical entry (e.g., "Ma" versus "Mom").

Discussion

Implications for Intelligibility Assessment in Conversational Speech

The primary purpose of this study was to determine whether a caregiver's gloss of her child with a speech delay might be useful when assessing intelligibility in conversational speech samples. Within the methodological constraints of this study, the findings suggest that caregivers do not

understand their children's speech as well as generally presumed. The fairly large individual differences among these caregivers' abilities to understand their children are consistent with data reported by Goehl and Martin (1987) for their three mother-child pairs. The ranges for words (42–98%) and utterances (48–98%) that caregivers thought they understood (i.e., glossed) suggest that there is a great deal of some children's speech that their primary caregivers acknowledge is unintelligible. Moreover, the statistically significant differences between what the caregivers thought they understood and the amount of speech they accurately understood, as assessed by intelligible words (26–96%) or fully intelligible utterances (29–95%), indicate that caregivers' glosses cannot necessarily be used as a veridical source of information about children's intended words.

The implications of these findings for the use of caregivers as informants for clinical or research intelligibility assessment protocols should be tempered by two methodological constraints in the present study. First, caregivers were not directly involved in the interaction with their child. Missing from their glossing task was the opportunity to manipulate topics, request repetition of utterances, or seek clarification as needed. These opportunities in dyadic communication might have provided the needed information for difficult glosses, as well as stimulated more focused attention on contextual cues available for all child utterances (cf. Monsen, 1983). Thus, although the present study simulated the typical clinical-research intelligibility assessment in which an examiner interacts with the child to obtain the speech sample, findings may underestimate caregivers' ability to gloss their child's conversational speech when they are the conversational partner.

The second methodological constraint in the present study is the level of responsibility placed on caregivers for generating glosses. The instructions did not explicitly request the caregiver to make every effort to gloss each word. Anecdotally, caregivers varied greatly in their investment in the task. Some appeared to make great efforts to gloss difficult sound sequences; others seemed content to guess or to quickly conclude that the sound sequence was hopelessly unintelligible. However, the findings for utterance accuracy compared to word accuracy suggest that caregivers may not have performed more accurately if pressed to gloss each word. The higher accuracy scores for utterance meaning might indicate that caregivers strive to grasp the general meaning of their children's utterances, rather than undertake word-for-word translation. As shown in Figure 3, caregivers were accurate in fully understanding only 58% of their child's utterances. However, when fully intelligible utterances were combined with those partially unintelligible utterances in which the meaning was retained (as identified in a comparison with the reference gloss), the caregivers accurately understood the meaning of 73% of their child's utterances (range: 35–98%). The finding that caregivers did not appear to use strategies to gloss unintelligible strings (such as trying out a word that would fit the syntactic context or begin with the first sound in the unintelligible string) suggests that caregivers' attention to meaning may impose a limit on word-for-word glossing. Specifically, the trend for caregivers to attribute more advanced rather than less advanced linguistic forms to their children may present a formidable methodological bias against using caregiver glosses. Thus, on a gradient of utility, caregivers' glosses may be most useful when supplying forms, such as proper nouns, and least useful when deliberating among precise inflectional and syntactic forms.

For each of the three purposes for intelligibility assessment reviewed at the outset of this paper, the present findings support the perspective that valid intelligibility assessment requires a sample of conversational speech. Interactive processing of segmental, suprasegmental, and language forms may underlie moments of unintelligibility, as suggested in the current study by the caregivers' ability to correctly gloss some strings and incorrectly gloss others that appeared to contain essentially similar phonologic targets and error types. This interaction among speech-language-prosody domains presumably is lost or seriously attenuated when children imitate words or sentences or respond in citation form speech. Yet knowledge of the words children intend to say during intelligibility breakdowns is needed for both the development of explanatory models and for intervention planning. Because the findings indicate that a significant percentage of caregivers' glosses of words is not reliable for these needs, a modification in the present methods is warranted.

One way to use a caregiver's gloss in intelligibility assessment might be to have both the caregiver and examiner independently gloss a videotape or audiotape of a child's conversational speech, with the caregiver encouraged to attend to each word. The examiner could then use both glosses to generate a narrow phonetic transcription of the tape, using specific knowledge of the child's speech errors and replays of the most difficult strings of unintelligible segments. The validity of words considered unintelligible in the examiner's original gloss, but glossed by the caregiver, could then be tested against the narrow phonetic transcription. Use of this obviously time-consuming procedure for all or certain parts of the conversational speech sample would at least constitute a controlled procedure for studies requiring close analyses of speech that is difficult to gloss.

Implications for Speech-Language Normalization and Clinical Management

The findings that caregivers have more difficulty than expected in understanding their children's speech and tend to overestimate their syntactic development when glossing difficult word strings raises questions about the impact of reduced intelligibility on these children's speech-language development and normalization. Emerging studies suggest that when children's utterances are unintelligible, caregivers are less likely to respond with facilitative language strategies, such as topic continuation and topic continuing prompts (Yoder & Davies, 1990), and simple recasts (Conti-Ramsden, 1990). Even in response to the intelligible utterances of a child with an intelligibility problem, caregivers may continue nonfacilitative conversational styles. For example, Gardner (1989) found that in contrast to caregivers of children with normally developing speech, caregivers of children with speech delays took more control of the conversational topic

(presumably as a strategy to increase intelligibility) when interacting with their child, thereby limiting opportunities for continuing a topic initiated by the child. Although differing perspectives exist on the function of caregivers' input for children's language development (cf. Furrow & Nelson, 1986; Gleitman, Newport, & Gleitman, 1984), training studies of children with normal language development (e.g., Cromer, 1987) and delayed language development (e.g., Culatta & Horn, 1982; Weismer & Murray-Branch, 1989) have demonstrated the effects of specific types of input on the acquisition of syntactic forms. Additional research describing exactly what speech-language forms caregivers impute to their inconsistently unintelligible children and how they temper their input accordingly could provide important information about normalization processes.

Most generally, the finding that caregivers have more difficulty than expected understanding their children's speech supports current clinical trends to make increased intelligibility with caregivers and others a central target in early intervention programs with children who have speech delays. Currently, the selection and sequencing of intervention targets to increase intelligibility are based on criteria such as types of phonological processes (Hodson & Paden, 1990) or frequency of occurrence of sounds in the language (Fudala, 1970). The differences in speech intelligibility among the 15 children in the current study, many of whom were using similar phonological processes and having difficulty with the same group of phonemes, suggest that generic target-selection guidelines for improving intelligibility may not address the relevant sources. What is lacking is a method for identifying the specific segmental errors that may contribute most to breakdowns in intelligibility for a specific child and the contributions of suprasegmental and language variables to those breakdowns. Studies of the interactive contributions of segmental, suprasegmental, and language variables to the moment-to-moment intelligibility of children's conversational speech could yield the information needed to develop rapidly effective intervention programs.

Acknowledgments

This research was funded by a grant from the National Institutes of Health, NIDCD, Grant No. DC00496. We thank Tammie Brochtrup, Cindy Campbell, Kirsten Curtis, Greta French, Karin Jones, Amy Knoeck, Lisa Kostka, Kris Krajewski, Susan Tachibana, Denise Timmer, and Lauren Walker for obtaining representative speech samples and providing the clinicians' glosses. We also thank the parents and children who participated in the study. Finally, special thanks to Dorothy Rorick for her editorial assistance and to Carol Stoel-Gammon, Shelley L. Velleman, and two anonymous *JSHR* reviewers for their cogent editorial suggestions.

References

Bernthal, J. E., & Bankson, N. W. (1988). *Articulation and phonological disorders.* Englewood Cliffs, NJ: Prentice Hall.

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4 year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52,* 156–173.

Camarata, S., & Leonard, L. B. (1986). Young children pronounce object words more accurately than action words. *Journal of Child Language, 13,* 51–65.

Campbell, T. F., & Shriberg, L. D. (1982). Associations among pragmatic functions, linguistic stress, and natural phonological processes in speech-delayed children. *Journal of Speech and Hearing Research, 25,* 547–553.

Conti-Ramsden, G. (1990). Maternal recasts and other contingent replies to language-impaired children. *Journal of Speech and Hearing Disorders, 55,* 262–274.

Cromer, R. F. (1987). Language growth with experience without feedback. *Journal of Psycholinguistic Research, 16,* 223–231.

Culatta, B., & Horn, D. (1982). A program for achieving generalization of grammatical rules to spontaneous discourse. *Journal of Speech and Hearing Disorders, 47,* 174–180.

Dyson, A. T., & Robinson, T. W. (1987). The effect of phonological analysis procedure on the selection of potential remediation targets. *Language, Speech, and Hearing Services in Schools, 18,* 364–377.

Fudala, J. B. (1970). *Arizona Articulation Proficiency Scale: Revised.* Los Angeles: Western Psychological Services.

Furrow, D., & Nelson, K. (1986). A further look at the motherese hypothesis: A reply to Gleitman, Newport & Gleitman. *Journal of Child Language, 13,* 163–176.

Gardner, H. (1989). An investigation of maternal interaction with phonologically disordered children as compared to two groups of normally developing children. *British Journal of Disorders of Communication, 24,* 41–59.

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language, 11,* 43–79.

Goehl, H., & Martin, P. (1987, November). *Intelligibility and familiarity: Is experience enough?* Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA.

Hodson, B. W., & Paden, E. (1981). Phonological processes which characterize unintelligible and intelligible speech in early childhood. *Journal of Speech and Hearing Disorders, 46,* 369–373.

Hodson, B. W., & Paden, E. (1990). *Targeting intelligible speech.* San Diego: College Hill.

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,* 482–499.

Klein, H. B. (1984). Procedure of maximizing phonological information from single-word responses. *Language, Speech, and Hearing Services in Schools, 15,* 267–273.

Klein, H. B., & Spector, C. C. (1985). Effect of syllable stress and serial position on error variability in polysyllabic productions of speech-delayed children. *Journal of Speech and Hearing Disorders, 50,* 391–402.

Miller, J. (1981). *Assessing language production in children.* Baltimore, MD: University Park Press.

Monsen, R. B. (1981). A usable test for the speech intelligibility of deaf talkers. *American Annals of the Deaf, 126,* 845–852.

Monsen, R. B. (1983). The oral speech intelligibility of hearing-impaired talkers. *Journal of Speech and Hearing Disorders, 48,* 286–296.

Morrison, J., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research 35,* 259–273.

Panagos, J. M., & Prelock, P. (1982). Phonological constraints on the sentence productions of language-disordered children. *Journal of Speech and Hearing Research, 25,* 171–177.

Panagos, J. M., Quine, M. E., & Klich, R. J. (1979). Syntactic and phonological influences on children's articulation. *Journal of Speech and Hearing Research, 22,* 841–848.

Prelock, P. A., & Panagos, J. M. (1989). The influence of processing mode on the sentence productions of language-disordered and normal children. *Clinical Linguistics & Phonetics, 3,* 251–263.

Prelock, P. A., & Panagos, J. M. (1991). Discourse-sentence interaction in the speech of language-disordered and normal children. *Clinical Linguistics & Phonetics, 5,* 79–88.

Schiavetti, N., Metz, D. E., & Sitler, R. W. (1981). Construct validity of direct magnitude estimation and interval scaling of speech intelligibility: Evidence from a study of the hearing impaired.

Journal of Speech and Hearing Research, 24, 441–445.

Shriberg, L. D. (1986). *PEPPER: Programs to Examine Phonetic and Phonologic Evaluation Records.* Hillsdale, NJ: Lawrence Erlbaum.

Shriberg, L. D. (in press). Four new speech and prosody measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research.*

Shriberg, L. D., & Kwiatkowski, J. (1980). *Natural Process Analysis (NPA): A procedure for phonological analysis of continuous speech samples.* New York: Macmillan.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders, 47,* 256–270.

Shriberg, L. D., & Kwiatkowski, J. (1985). Continuous speech sampling for phonologic analyses of speech-delayed children. *Journal of Speech and Hearing Disorders, 50,* 323–334.

Shriberg, L. D., & Kwiatkowski, J. (1988). A follow-up study of children with phonologic disorders of unknown origin. *Journal of Speech and Hearing Disorders, 53,* 144–155.

Shriberg, L. D., Kwiatkowski, J., Best, S., Hengst, J., & Terselic-Weber, B. (1986). Characteristics of children with speech delays of unknown origin. *Journal of Speech and Hearing Disorders, 51,* 140–161.

Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C. (1990). *The Prosody-Voice Screening Profile (PVSP).* Tucson, AZ: Communication Skill Builders.

Shriberg, L. D., Kwiatkowski, J., & Snyder, T. (1989). Tabletop versus microcomputer-assisted speech services: Stabilization phase. *Journal of Speech and Hearing Disorders, 54,* 233–248.

Weismer, S. E., & Murray-Branch, J. (1989). Modeling versus modeling plus evoked production training: A comparison of two language intervention methods. *Journal of Speech and Hearing Disorders, 54,* 269–281.

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

Weston, A. D., & Shriberg, L. D. (in press). Contextual and linguistic correlates of intelligibility in children with developmental phonological disorders. *Journal of Speech and Hearing Research.*

Yoder, P. J., & Davies, B. (1990). Do parental questions and topic continuations elicit replies from developmentally delayed children? A sequential analysis. *Journal of Speech and Hearing Research, 33,* 563–573.

Yorkston, K. M., & Beukelman, D. R. (1981). *Assessment of intelligibility of dysarthric speech.* Tigard, OR: C. C. Publications.

Received June 3, 1991
Accepted January 31, 1992

Contact authors: Joan Kwiatkowski or Lawrence D. Shriberg, Phonology Project, The Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, Wisconsin 53705.

Intelligibility Assessment in Developmental Phonological Disorders: Accuracy of Caregiver Gloss

Joan Kwiatkowski, and Lawrence D. Shriberg
J Speech Hear Res 1992;35;1095-1104

This information is current as of June 28, 2012

This article, along with updated information and services, is
located on the World Wide Web at:
http://jslhr.asha.org/cgi/content/abstract/35/5/1095

AMERICAN
SPEECH-LANGUAGE-
HEARING
ASSOCIATION