Journal of Speech and Hearing Research, Volume 35, 259–273, April 1992

Articulation Testing Versus Conversational Speech Sampling

Judith A. Morrison
University of Redlands
Redlands, CA

Lawrence D. Shriberg
University of Wisconsin-Madison

Detailed speech analyses were performed on data from 61 speech-delayed children assessed by both a standard articulation test and a conversational speech sample. Statistically significant differences between the articulation accuracy profiles obtained from the two sampling modes were observed at all linguistic levels examined, including overall accuracy, phonological processes, individual phonemes, manner features, error-type, word position, and allophones. Established sounds were often produced more accurately in conversational speech, whereas emerging sounds were often produced more accurately in response to articulation test stimuli. Error patterns involving word-to-word transitions were available only in the context of continuous speech. A pass-fail analysis indicated that the average subject would receive similar clinical decisions from articulation testing and conversational speech sampling on an average of 71% of consonant sounds. Analyses of demographic, language, and speech variables did not yield any subject characteristics that were significantly associated with concordance rates in the two sampling modes. Discussion considers sources of variance for differences between sampling modes, including processes associated with both the speaker and the transcriber. In comparison to the validity of conversational speech samples for integrated speech, language, and prosodic analyses, articulation tests appear to yield neither typical nor optimal measures of speech performance.

KEY WORDS: phonology, assessment, speech sampling, continuous speech, articulation tests

The term *communicative competence* refers to performance in everyday social interaction (Hymes, 1971). It is only in such contexts that we can observe the variety of human discourse and the complex speech-language interactions that occur when one is talking to be understood. Despite agreement on the face validity of such contexts for assessing what might be termed *typical* or *customary* communicative performance (e.g., Miller, 1981; Locke, 1983; Wren, 1985), clinicians and researchers continue to justify the use of alternative assessment procedures on the basis of efficiency or utility considerations.

Nowhere in the communicative disorders literature are efficiency issues more evident than in the selection of sampling procedures for phonological analysis. Particularly in child phonology, single word articulation tests continue to be widely used, despite over four decades of research documenting differences that occur when children give single-word responses, termed *citing,* compared to their speech when spontaneously *talking* (e.g., Andrews & Fey, 1986; Dubois & Bernthal, 1978; Dyson & Robinson, 1987; Faircloth & Faircloth, 1970; Gallagher & Shriner, 1975; Harris & Cottam, 1985; Healy & Madison, 1987; Johnson, Winney, & Pederson, 1980; Jordan, 1960; Kenney, Prather, Mooney, & Jeruzal, 1984; Klein, 1984; Klein & Spector, 1985; Orr, Blodgett, & Miller, 1983; Paden & Moss, 1985; Paynter & Sims, 1979; Schmitt, Howard, & Schmitt, 1983; Siegel, Winitz, & Conkey, 1963; Simmons, Blodgett, & Miller, 1983; Van Demark, 1964). Unlike case studies or small group studies of normal speech development that have used continuous speech samples collected in natural

0022-4685/92/3502-0259$01.00/0

or experimental settings (e.g., Leopold, 1947; Stemberger, 1988; Stoel-Gammon, 1987; Vihman & Greenlee, 1987), large-scale studies of both normal and speech-delayed children most often have been based on diverse citation-form tasks and protocols, most frequently in the form of pictured-word articulation tests.

A review of over 50 unpublished and published studies during the past 40 years indicates a diversity of descriptive methods to assess the clinical consequences of different forms of citation-form testing or to compare citation-form testing with continuous-speech sampling. Dependent variables in this literature parallel the chronology of theoretical paradigms in developmental speech disorders. More recent studies have focused on the potential effects of sampling mode on the frequency of occurrence of errors classified by phonological processes. Table 1 provides a descriptive summary of 20 representative studies reported during the past two decades. These studies were selected to illustrate the methodologic diversity from which generalizations about the influence of type and mode of sampling have been based. Sample sizes range from 1 to 240, with children's ages ranging from under 3 years to over 12 years. Most studies using citation forms have been based on spontaneous rather than imitative responses, with 10 different articulation tests represented within just this sample of the literature. Thus a major constraint on generalizations from the citation-form data across studies is that they reflect unique (i.e., unreplicated) permutations of modes of evocation with specific pictured-word stimuli. Even the size and color of materials used to evoke the same pictured words have been studied as potential sources of variance (e.g., Bernthal, Grossman, & Goll, 1989). Of three types of contexts used to sample continuous speech—spontaneous, retelling, and imitation—most studies have used spontaneous conversational speech. However, within the studies using continuous speech samples, considerable differences in method are apparent, including differences in the child's purpose for talking (e.g., naming, repeating, informing), the level of propositional effort (e.g., labelling, telling, retelling, describing), the availability of situational support for talking (e.g., familiar or observable objects, pictures, or events), the level of spontaneity reflected in the sample (e.g., spontaneous, evoked, imitative), the length and complexity of the utterance, and differences in the semantic, syntactic, and phonetic content of the comparison articulation test.

This diversity of methodological approaches notwithstand-

TABLE 1. A representative sample of 20 studies (1970–1990) illustrating the methodological diversity in research comparing phonological performance in citation forms and continuous speech.

	Subjects			Mode of evocation and sampling context				
				Citation forms		Continuous speech		
Study	*n*	Age (Yrs:Mos)	Speech Status	Spontaneous	Imitated	Spontaneous	Retelling	Imitated
Faircloth & Faircloth (1970)	1	11:0	Speech-delayed		x[a]	x		
Chapman & Ting (1971)	40	Preschool–1st Grade	Normal	x[a]	x[a]			x[a]
DuBois & Bernthal (1978)	18	4:3–6:2	Speech-delayed	x[b]	x[c]	x	x	
Paynter & Sims (1979)	4	4:11–5:11	Language-delayed	x[d]	x[c]	x		
Johnson, Winney, & Pederson (1980)	35	3:7–9:5	Speech-delayed	x[e]			x[e]	
Shriberg & Kwiatkowski (1980)	10	3:0–6:0	Speech-delayed	x[f]	x[f]	x		
Dunn (1982)	1	4:6	Speech-delayed	x[g]	x[h]	x		
Bankson & Bernthal (1982)	18	4:0–4:11	Speech-delayed		x[h]			x[f]
Schmidt, Howard, & Schmidt (1983)	240	3:0–7:0	Normal	x[i]		x		
Simmons, Blodgett, & Miller (1983)	8	3:9–4:6	Speech-delayed	x[j]		x		
Klein (1984)	10	4:1–6:1	Speech-delayed	x[f]		x		
Klein & Spector (1985)	8	5:2–6:11	Speech-delayed	x[b]		x		
Paden & Moss (1985)	3	4:11–7:6	Speech-delayed	x[g]		x		
Andrews & Fey (1986)	14	2:8–6:1	Speech-delayed	x[g]		x		
Blodgett & Miller (1986)	15	3:1–4:9	Speech-delayed	x[a]		x		x[a]
Healy & Madison (1987)	20	5:4–12:8	Speech-delayed	x[k]		x[c]		
Dyson & Robinson (1987)	5	3:5–6:5	Speech-delayed	x[g]		x		
Watson (1989)	8	3:1–7:4	Speech-delayed	x[l]	x[m]	x		
Elbert, Dinnsen, Swartzlander, & Chin (1990)	10	3:7–5:9	Speech-delayed	x[b]		x		
Smit (1990)	21	3:0–39:9	Normal					x[n]

Note. [a] Same words taken from spontaneous conversational speech; [b] Authors' word list; [c] Examiner elicited the desired response if not produced spontaneously; [d] Templin-Darley Test of Articulation (Templin & Darley, 1969); [e] Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 1972); [f] Photo Articulation Test (Pendergast, Dickey, Selmar, & Soder, 1969); [g] Assessment of Phonological Processes (Hodson, 1980); [h] Phonological Process Analysis (Weiner, 1979); [i] Arizona Articulation Proficiency Scale (Fudala, 1974); [j] Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 1969); [k] Weiss Comprehensive Articulation Test (Weiss, 1978); [l] Compton-Hutton Phonological Assessment (Compton & Hutton, 1978); [m] Assessment Link Between Phonology and Articulation (Lowe, 1986); [n] Tape-recorded utterances with rate variations.

ing, some common trends in the literature on the effects of speech sampling mode on speech data are apparent. Table 2 is a summary of effects that have received some support across several studies. In general, more frequent and varied errors occur in talking compared to citing, but group and individual-level findings also include instances in which speech errors are more frequent in citation modes. The more recent studies using the phonological process as the unit of articulatory analysis have the disadvantage of concealing specific descriptive information on the interaction of error sounds by error types by word positions (cf., Shriberg, 1990). Most generally, continuous speech appears to be associated with more deletion errors, especially of consonants in word-final position, and with increased errors involving clusters and unstressed syllables. In recognition of these frequently observed yet unpredictable differences, most current assessment guidelines recommend using narrowly transcribed, whole-word analyses of both citation-form and continuous speech sampling. Sampling in multiple contexts presumably ensures the broadest data base from which to make judgments about what Stoel-Gammon (1988) has termed children's *relational* and *independent* phonological competence (Bernthal & Bankson, 1988).

Explanations for the findings of more articulation errors in either talking or citing (see Table 2) have focused on three potential sources of variance. Most salient are the consequences of differences in the speech sample's linguistic structure and content, with its associated demands on phonological and speech-motor processing. Linguistically, the canonical complexity of stimulus words in articulation tests has been viewed as increasing the probability of phoneme deletions of consonant singletons, consonant clusters, and unstressed syllables. More generally, the pattern of errors obtained from a speech sample might be a reflection of the number, complexity, and saliency of multisyllabic forms in the corpus (e.g., Klein & Spector, 1985; Shriberg & Kwiatkowski, 1980; Simmons, Blodgett, & Miller, 1983). The acquisition of continuous speech rules at the phonological level involves complex interactions among semantic, syntactic, morphologic, pragmatic, and prosodic tiers (Levelt, 1989; Selkirk, 1984). Such structural and phonological complexities are presumed to invoke resource allocation processes, as observed in both younger normal-speaking children (e.g., Dunn & Davis, 1983) and adults with other communicative disorders (e.g., Kohn, 1988).

A second potential source of explanation for group- and individual-level differences in error sound patterns in the two sampling modes is the contextual support available to the speaker in talking compared to citing. Although resource allocation demands presumably are greater in talking, conversational discourse may offer a different form of cognitive-linguistic support than available during articulatory testing. Snyder's (1984) discussion of children's communicative competence reflects such a perspective, including observations on how such information might contribute to estimates of a child's typical communicative ability as well as predict a child's potential for communicative growth:

> . . . if we can develop a model of communicative competence that can accommodate the diversity of communicative output, its processing, and the competitive and collaborative use of resources within the system, we might make more realistic assumptions and predictions about children's communicative

TABLE 2. Some literature findings comparing phonological performance in citation forms (citing) and continuous speech (talking).

Construct	Findings	References
Severity of involvement	More articulatory errors occur in talking than citing, but severity ratings may be equivalent or poorer when based on citation forms.	Andrews & Fey, 1986; DuBois & Bernthal, 1978; Faircloth & Faircloth, 1970; Healy & Madison, 1987; Johnson, Winney, & Pederson, 1980; Klein, 1984; Simmons, Blodgett, & Miller, 1983
Error types	The majority of error types using phonological process typologies have been observed in both talking and citing: Assimilation; Cluster Reduction; Gliding: Liquid Deviation/Simplification; Medial Consonant Deletion; Palatal Deviation/Fronting; Stopping; Vocalization.	DuBois & Bernthal, 1978; Kenney, Prather, Mooney, & Jeruzal, 1984; Klein, 1984; Paden & Moss, 1985; Simmons et al., 1983
	Certain error types are more frequent in talking than citing: Cluster Reduction; Consonant Deletion, Final Consonant Deletion; Syllable Deletion, Unstressed Syllable Deletion. Substitution errors in citing may be realized as deletions in talking.	Andrews & Fey, 1986; DuBois & Bernthal, 1978; Faircloth & Faircloth, 1970; Healy & Madison, 1987; Johnson et al., 1980; Klein, 1984; Orr, Blodgett, & Miller, 1983; Paynter & Simms, 1979
Individual differences	Several subjects in several studies have demonstrated errors on several sounds or error classes only in talking contexts: Ambisyllabic Consonant Deletion; Assimilation; Coalescence; Final Consonant Deletion; Initial Consonant Deletion; Neutralization; Stopping; Unstressed/Weak Syllable Deletion.	DuBois & Bernthal, 1978; Dyson & Robinson, 1987; Healy & Madison, 1987; Johnson et al., 1980; Klein, 1984; Orr et al., 1983; Paynter & Sims, 1979
	Several subjects in several studies have demonstrated errors on several sounds and error classes only in citing: Glide and /l/ Deviation; Labial Assimilation; Stridency Deletion; Velar Deviation.	Andrews & Fey, 1986; Dyson & Robinson, 1987; Paden & Moss, 1985

> competence. If we consider the subject and task characteristics as well as the resources that can be used to perform various communicative tasks, we may be able to unmask some of the puzzling inconsistency in the language-delayed child's communicative performance (p. 104).

Snyder's perspective suggests that spontaneous conversational speech may not represent a more difficult context than citation forms due to a child's opportunity to partially control topic and content, thus ensuring familiarity of words and meanings. Findings by Menyuk (1980) and Campbell and Shriberg (1982) demonstrate variability in phonological accuracy associated with linguistic stress and pragmatic function, with improved accuracy when children assert or clarify messages in free conversation. Whereas error frequency may increase in longer, phonologically more complex words, unstressed monosyllabic words in conversation (which often carry less communicative force) appear to be most vulnerable to articulation errors (Klein & Spector, 1981).

A third and seldom discussed source of variance in speech sampling reflects the effects of speaker and task variables on phonetic transcription. Response definitions for acceptable articulation in citation forms typically differ from those required to assess articulation in continuous speech forms. The latter require familiarity with many assimilatory processes that operate in casual and fast speech. For example, if asked to read the sentence "He put his hat on," normal adult and child speakers would articulate each of the word-initial /h/ sounds correctly. In spontaneous conversational speech, however, whether from an adult or child with normal or disordered speech, deletion of the initial /h/ in the unstressed pronoun "his" is common and acceptable. The contributions of such linguistic variables in continuous speech to speech perception outcomes in normal discourse have received considerable research attention (Klatt, 1989; Nittrouer & Boothroyd, 1990). Alternatively, over-careful articulation of sounds in citation forms (e.g., adding a slight schwa off-glide to word final voice stops) would generally be recorded as "correct" unless the examiner had reason to believe that the speaker was not in control of the appropriate allophone (e.g., as might be the case for a speaker in an accent reduction program). Thus, whether using correct-incorrect scoring, broad phonetic transcription, or narrow phonetic transcription, measurement decisions must be guided by explicit response definitions for each sampling mode.

This paper proceeds from the positions advanced by Henderson (1938) and Jordan (1960), which claim that the appropriate referent for the concurrent validity of articulation tests is the construct of *customary* speech performance in contextually rich continuous speech. Drawing from a database of diagnostic assessments of young, speech-delayed children, we examined children's responses to citation-form articulation testing to assess in some detail whether such responses reflect their performance in spontaneous conversational speech.

Method

Subjects

Transcripts from 61 of a cohort of 64 speech-delayed children in a longitudinal study met criteria for inclusion in the study. The 64 children had been referred by speech-language pathologists for intelligibility deficits of unknown origin. Only those children who produced at least 50 nonimitative utterances in each of two sampling conditions (to be described) were included in the present analyses. As shown in Table 3, the gender ratio favoring boys by approximately 2:1 was lower than the approximately 3:1 ratio reported in previous studies of such children (Shriberg, Kwiatkowski, Best, Hengst, & Terselic-Weber, 1986). Consistent with prior work, approximately 85% of subjects were 4–6 years of age. A metric termed Percentage of Consonants Correct (PCC) (Shriberg & Kwiatkowski, 1982; Shriberg et al., 1986), calculated on the conversational speech samples used for this study, classified most children's speech delays as falling in the mild-moderate to moderate-severe range. All children had hearing within normal limits, showed no significant sensory-motor involvement, and were developing normally in cognitive and social domains. With the exception of one child, language comprehension was within normal limits, and the children's expressive language ranged from within normal limits to moderately delayed. The assessment protocol (see below) included both standardized measures of language comprehension and production and free speech sampling procedures (Miller, 1981).

Procedures

All data collection and analyses procedures used in the present study have been developed and described in prior work, including procedures for sampling speech, accomplishing narrow phonetic transcription by consensus, and coding and entering transcriptions for computer-aided phonological analysis (Shriberg, 1986; Shriberg & Kent, 1982; Shriberg & Kwiatkowski, 1980, 1985; Shriberg et al., 1986; Shriberg, Kwiatkowski, & Hoffmann, 1984; Shriberg, Hinke, & Trost-Steffen, 1987). Audiocassette samples of conversational

TABLE 3. Description of subjects.

			Age (years)				Percentage of Consonants Correct (PCC)[a]		
Gender	*n*	%	3	4	5	6	Mean	Standard Deviation	Range
Male	40	66	6	18	14	2	61.5	6.9	47.1–74.6
Female	21	34	3	1	6	1	64.5	10.0	38.9–78.3
Both	61	100	9	29	20	3	62.5	8.2	38.9–78.3

Note. [a] Shriberg & Kwiatkowski (1982).

speech and responses to the Photo Articulation Test (Pendergast, Dickey, Selmar, & Soder, 1969) were obtained in randomized order by trained research examiners as part of a 2½-hour diagnostic assessment protocol. Specifically, the ordering of the articulation test and the conversational speech sample depended on the examiner's judgment on how comfortable each child appeared with the more structured articulation test compared to the less structured conversational sample. All testing was done in a quiet suite using high-quality audiocassette tape on Marantz PMD221 audiocassette recorders, with matching external microphones monitored at a lip-to-microphone distance of approximately 15cm. Responses to the articulation test were obtained spontaneously using the set of picture cards and evocation procedures recommended for this standardized, single-word, citation-form measure. Conversational samples were obtained by inviting the child to talk about home and social activities, using a set of cues and prompts described in Shriberg and Kwiatkowski (1985).

Narrow phonetic transcriptions of all samples were completed by two two-person consensus teams that were well trained in a set of computer-aided transcription methodologies for young, speech-disordered children (Shriberg, 1986; Shriberg & Kent, 1982; Shriberg & Kwiatkowski, 1980). Using well-maintained Dictaphone 2025 play-back devices, one team transcribed the continuous speech samples, and the other transcribed each of the words used as articulation test stimuli. That is, for all analyses of the citation forms to follow, computations are based on all occurrences of vowel/diphthong and consonant sounds that occurred in each of the 76 articulation test words. A 90–70–225 sampling rule ensured comparable continuous speech samples containing either 90 word types, 70 utterances, or a total of 225 words, whichever criterion was met first during the glossing-transcription process.

Reliability

Extensive interjudge and intrajudge reliability data for each of the two-person consensus transcription teams used in this study have been reported, including individual sound-level data from six separate studies of children and adults with speech disorders (Shriberg & Lof, 1991). Extremely conservative agreement criteria were used in a computer program that calculated transcription reliability (Shriberg & Olson, 1987), with narrow phonetic transcription agreement on a sound requiring exactly the same configuration of any of the 42 diacritics available to describe a speech error. Using randomly selected transcripts, including approximately 100 to over 250 tokens per comparison, interjudge agreement between the two teams on articulation test and conversational speech samples ranged from 61.3% to 70.1% for narrow phonetic transcription of consonants and vowels/diphthongs and 84.7% to 88.6% for broad transcription of consonants and vowels/diphthongs. Intrajudge agreement for each team, calculated on approximately 80 to 1,100 tokens per comparison, was 65.5% to 81.1% for narrow phonetic transcription and 86.7% to 95.1% for broad phonetic transcription. These transcription agreement figures are consistent with other reports in disordered child phonology, with narrow phonetic transcription reliable for only certain research questions and broad phonetic transcription coefficients achieving adequate interjudge and intrajudge levels (cf., Shriberg & Lof, 1991). As discussed in Shriberg and Lof, transcription agreement on continuous speech samples appears to be somewhat higher than agreement based on articulation test responses (1–13 percentage points) due to complex associations among word forms and error types in each mode. The following findings, with the exception of the allophone data, are based on analyses of the broad phonetic transcriptions.

Structural Comparison

Previous analyses indicate that the types and percentages of word forms and phoneme distributions differ in continuous speech compared to articulation test protocols (Shriberg, 1986; Shriberg & Kwiatkowski, 1980). Before proceeding to the analyses, it is important to compare the distributions of intended (i.e., correct adult) canonical forms and the distributions of intended consonants and vowels/diphthongs in the present two sampling modes. The four panels in Figure 1 are graphic summaries of the relevant data. The top left panel is a display of the percentage of intended canonical forms in the two sampling modes, with intended percentage of occurrence of 10 forms sorted left to right from most to least frequent in conversational speech. Wilcoxon Matched Pairs Signed-Ranks tests indicated that with the exception of the percentage of intended CVC forms, all comparisons were significant at the .0001 level of confidence. The largest apparent absolute differences are in the average percentages of intended CV, two-syllable, and VC words in each mode (there were no intended V or CnV forms in the articulation test), with the most notable difference in the percentage of two-syllable words in continuous speech compared to articulation tests. A likely source of these differences is the lack of function words in articulation test stimuli. Simple word shapes, such as those for some determiners, are underepresented in standardized articulation test stimuli, whereas they account for over 22% of words in comparable continuous speech samples (Shriberg & Kwiatkowski, 1985). Moreover, the need to test sounds in the intervocalic position in articulation tests is associated with the greater number of two- and three-syllable words in the articulation test, compared to their average occurrence in the continuous speech samples. The general picture across the 10 categories is that in comparison to word forms in conversational speech, those in this articulation test present considerably more difficult structural contexts for articulatory performance.

The remaining three panels in Figure 1 provide sampling mode comparisons for the percentage of occurrence of intended vowels/diphthongs, singleton consonants, and consonant clusters. The continuous speech data points are again sorted left to right from most to least frequent. The trends suggest that relatively large percentage differences occurred for only a few sounds in each class. For vowels/diphthongs, the greatest differences in the percentage of intended sounds in each sampling mode occurred on the

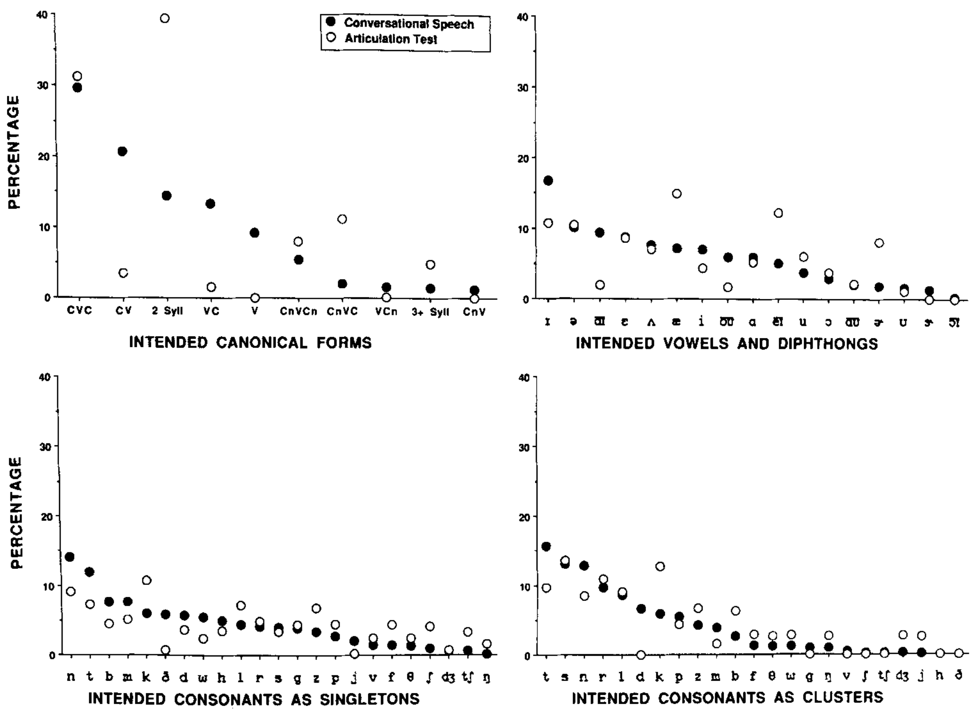

FIGURE 1. Percentage intended canonical and segmental forms in conversational speech and articulation testing sorted left to right from most to least frequent in conversational speech.

three diphthongs /aɪ/, /oʊ/, and /eɪ/ and the two vowels /æ/ and /ɚ/. These differences can readily be accounted for by both lexical differences in the two sampling modes (e.g., frequent use in conversational speech of /aɪ/, "I") and differences associated with lexical and phrasal destressing in conversational speech. Trends were generally similar for the consonant singleton and consonant cluster data, with few relatively large differences in the intended occurrence of sounds (e.g., more frequent intended /ð/ in demonstratives in conversational speech) and some intended vowel and diphthong sounds in conversational speech not occurring in the articulation test stimuli used for this study.

Results

The nonparametric Wilcoxon Matched Pairs Signed-Ranks test was used as the advisory inferential statistic for the major between-modes analyses. Although distributional considerations required the use of nonparametric statistics (including the large number of 0% and 100% scores that could not be adjusted using transformations such as the arcsin), parametric descriptive statistics are used in the following figures to best illustrate patterns of central tendency and dispersion.

Overall Accuracy

Figure 2 is a display of the overall accuracy of subjects in each mode, with means and standard deviations for percentages of correct vowels/diphthongs, consonant singletons, and consonant clusters shown for each of the two sampling modes. The stressed vowel /ɝ/ and the diphthong /ɔɪ/ comparisons were excluded from these analyses; as indicated in Figure 1, these sounds are not included in the pictured-word section of the articulation test. The overall accuracy of sound production in both conditions followed predictable developmental patterns. These speech-delayed children produced vowels and diphthongs most accurately, with substantially lowered accuracy and greater interchild variability on consonant singletons and clusters. As shown by the asterisks, vowels/diphthongs and consonant singletons were significantly more accurate in continuous conversational speech, a finding that is not consistent with trends reported in the speech sampling literature. Analysis at the level of individuals confirmed these significant group-level findings for the consonant data. Specifically, the results of a Percentage of

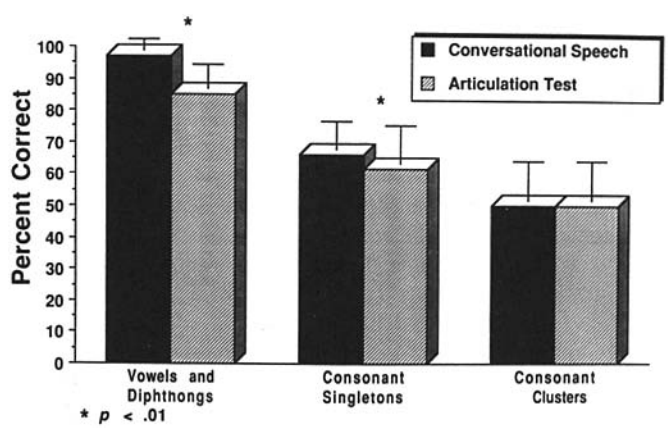

FIGURE 2. Overall articulatory accuracy in conversational speech and articulation testing. Significant differences were tested using Wilcoxon Matched Pairs Signed-Ranks Test.

Consonants Correct calculation on the consonant data in both modes indicated that 47 of the 61 subjects (77%) had better articulation in the conversational speech mode. Description of the specific group- and individual-level sources of these differences in overall accuracy are explored in the following analyses.

Natural Phonological Processes Analysis

Natural processes analyses provide a summary of errors aggregated in descriptive units that, as noted previously, reflect various combinations of error types on certain sounds in certain word positions. Results of such analyses, using the definitions for eight natural processes described in Shriberg (1986) and Shriberg and Kwiatkowski (1980), are shown in Figure 3. The average percentage of occurrence of each of the eight processes in each sampling condition are plotted in descending order based on the data for conversational speech samples. As shown in Figure 3, the general shape of the trends for process occurrence in the two modes were similar. However, statistically significant differences at the .01 level or greater were obtained on six of the eight Wilcoxon Matched Pair Signed-Ranks test comparisons. Recall that in Figure 2, as well as in the figures to follow, the group means are plotted, whereas the nonparametric statistical tests reflect individual rank-order differences in performance data obtained in each sampling mode. Hence, some of the data points will appear to be inconsistent with the statistical findings indicated by the asterisks.

Consistent with previous studies (Dunn & Davis, 1983; Klein, 1984; Dyson & Robinson, 1987), Cluster Reduction and Liquid Simplification were the most frequent process-level error descriptors in both sampling conditions. In contrast to prior reports, however, these descriptive categories significantly more often characterized articulation test responses. Moreover, as shown in Figure 3, both Stopping and Final Consonant Deletion occurred significantly less frequently in response to articulation testing, compared with their occurrence in conversational speech. Just over half of these subjects (32 of 61) had greater than 20% occurrence of Stopping in their continuous speech samples, but no occurrences of Stopping in the articulation test data. Specifically, most children stopped initial fricatives significantly more often in conversational speech, and over 50% of these children demonstrated significantly higher rates of final /t/ deletion in conversational speech. Finally, citation form testing yielded significantly higher frequencies of occurrence for the least frequent of the eight processes, as indicated by data points below 10% for Unstressed Syllable Deletion and Assimilation.

Segmental Analysis

Vowels and diphthongs. Figure 4 includes the descriptive data and results of inferential statistical testing for the percentage of correct vowels/diphthongs in the two sampling modes. Although the majority of vowels/diphthongs were highly accurate in both samples, significant differences between samples were obtained for four sounds. The vowels /ə/ and /ɪ/ were produced more accurately in continuous speech, whereas /aɪ/ and /ɚ/ were more accurate in articulation testing. Of these four differences, the findings for /ə/ were least expected, yet most robust. The continuous speech samples included a number and variety of lexical items containing this reduced vowel form, with its occurrence in the articulation test words limited to the subset of multisyllabic words described above in the analyses of structural differences between sampling modes. In articulation testing, children more often deleted /ə/ in the unstressed syllable of the multisyllabic words or replaced /ə/ with another stressed vowel. In contrast, for the relatively lower proportion of multisyllabic words in continuous speech, /ə/ was more often preserved as unstressed. Moreover, /ə/ was used often as an acceptable unstressed vowel in casual speech forms.

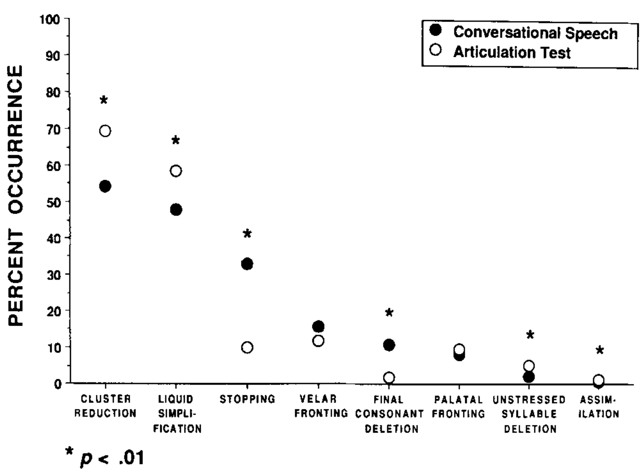

FIGURE 3. Natural phonological process occurrence in conversational speech and articulation testing. Mean percent occurrences is sorted left to right from most to least frequent in conversational speech. Significant differences were tested using Wilcoxon Matched Pairs Signed-Ranks Test.

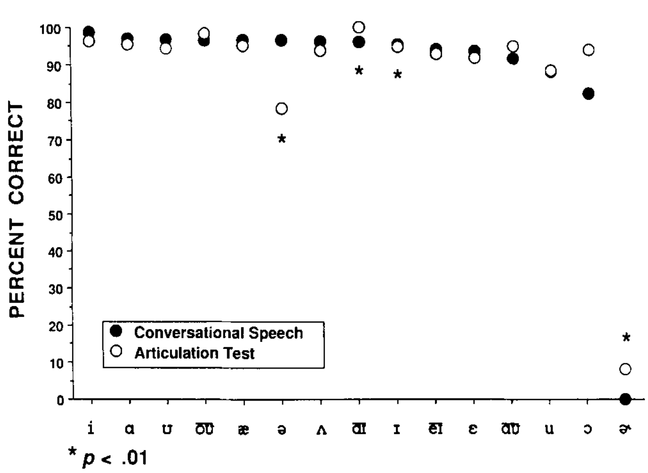

FIGURE 4. Mean percentage of correct vowels and diphthongs in conversational speech and articulation testing sorted left to right from most to least frequent in conversational speech. Significant differences were tested using Wilcoxon Matched Pairs Signed-Ranks Test.

Stress appeared to be implicated in the other vowel differences as well. For example, the unstressed vowel /ɚ/ was rarely correct in continuous speech (mean = 0%; range = 0%–25% accuracy), but it averaged approximately 8% accuracy in articulation test responses. This difficult sound for speech-delayed children was produced with somewhat greater accuracy in the context of the word stress associated with articulation testing (i.e., isolated, citation-form responses). Although the stressed rhotic vowel /ɝ/ occurred with insufficient frequency in articulation test forms to allow for statistical comparison, children also produced /ɝ/ correctly more often in articulation testing, thus adding additional support to the association of stress with articulation accuracy.

Consonants and features. Figure 5 displays the group means for individual consonants in the two sampling modes. The dashed line divides the 23 consonants into those averaging greater and less than 50% correct in the two sampling modes. As shown by the trends in Figure 5, most sounds averaging greater than 50% accuracy were more accurate when sampled in conversational speech, whereas those averaging less than 50% correct were more correct when sampled by articulation testing. Of the former group, average percentage correct differences were statistically significant at the .01 level or greater for /j/, /m/, and /n/; of the latter group, average percentage correct differences were statistically significant for /ʃ/, /s/, /z/, /l/, and /r/. The one exception to this trend was /g/. Although /g/ was produced with greater than 50% accuracy in both samples, it tended to be more accurate in articulation test responses.

Figure 6 displays the consonant data shown in Figure 5, as aggregated by six manner features. Nasals, glides, and stops were predictably more accurate in both sampling conditions compared to affricates, fricatives, and liquids. However, as indicated in Figure 6, average performance within these trends on nasals and glides was significantly better in conversational speech compared to articulation testing; average performance on affricates and fricatives was significantly

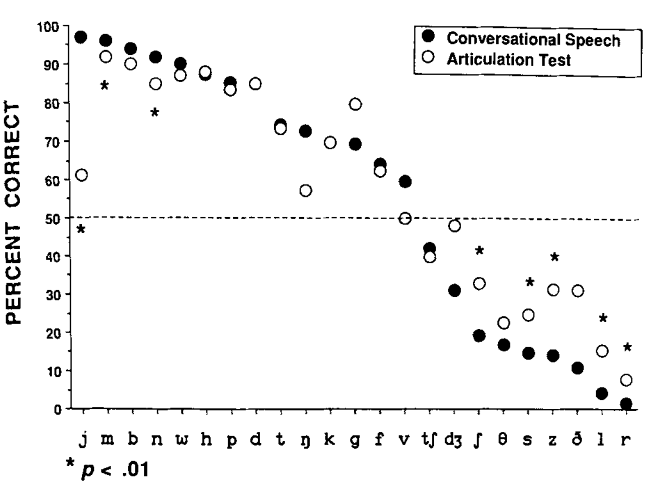

FIGURE 5. Mean percentage of correct consonants in conversational speech and articulation testing sorted left to right from most to least frequent in conversational speech. Significant differences were tested using Wilcoxon Matched Pairs Signed-Ranks Test.

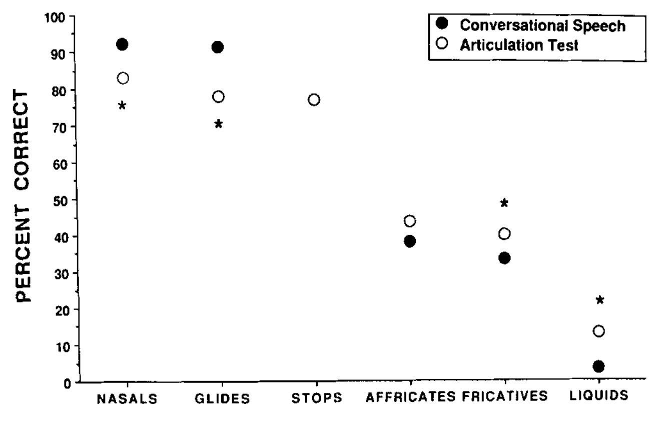

FIGURE 6. Mean percentage of correct segments by manner features in conversational speech and articulation testing sorted left to right from most to least frequent in conversational speech. Significant differences were tested using Wilcoxon Matched Pairs Signed-Ranks Test.

better in articulation testing. The reduced number of pairwise contrasts for the affricates (14 children did not attempt /tʃ/ in conversational speech; 17 did not attempt /dʒ/) lowered the statistical power for these comparisons.

Error type and word position. Two additional segmental analyses compared the types of consonant errors made in each word position. The six panels in Figure 7 display the percentage of deletions, substitutions, and distortions in word-initial, word-medial, and word-final positions separately for consonant singletons and consonant clusters. In word-initial position, the error-type trends for each sampling mode are extremely similar to one other, with considerably more substitutions occurring on singleton consonants in both modes and relatively equal proportions of each error type occurring in initial clusters. In contrast, in word-medial position, the interactions of error-type and singleton-cluster with sampling mode are too complex for ready summary. In word-final position, the trends for singletons and clusters are essentially similar: proportionally more deletion and distortion errors occur in continuous speech, but more frequent substitution errors occur in articulation testing.

For an additional perspective on error types at the level of individual subjects, relative error-type percentages were calculated for each of the three error types. For this calculation, a subject's total number of consonant errors was used as the denominator and numbers of each error type as numerators, thus controlling for subjects' severity of involvement. The three panels in Figure 8 provide subjects' relative error scores in each sampling mode for deletions, substitutions, and distortions. These subject data clearly indicate the trend for children's errors to be transcribed proportionally more often as substitutions when sampled from the articulation tests. Specifically, over 90% of the children had a higher percentage of relative substitution errors in response to the articulation test stimuli, compared to the percentage of relative substitution errors obtained in spontaneous continuous speech. In comparison, relative deletion and distortion errors were more nearly equal across sampling modes: 39% of

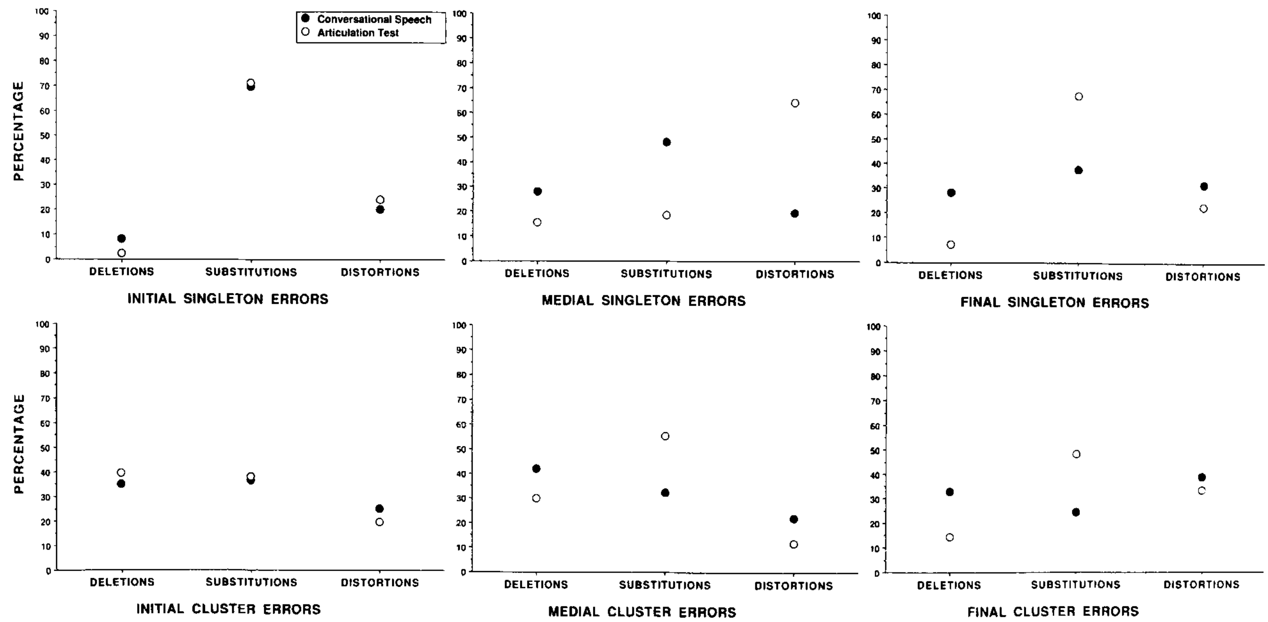

FIGURE 7. Error-type analyses of consonant singletons and consonant clusters in conversational speech and articulation testing.

children had a higher percentage of relative deletion errors, and 46% of children had a higher percentage of relative distortion errors in response to the articulation test stimuli.

Subphonemic Analysis

Narrow phonetic transcription allowed inspection of allophone-level differences in children's productions of vowels/diphthongs and consonants. These changes included the percentage of occurrence of casual and fast speech changes, as well as clinically relevant sound distortions and additions. Figure 9 shows the average percentage of occurrence of 34 diacritic symbols in the transcripts from the two speech sampling modes. Nearly half of the advisory rank-order statistical comparisons were significant at or beyond the .01 level; recall that the data points in these figures reflect means data. The pattern of diacritic use transcribed for the articulation test responses in Figure 9 can be characterized as showing strengthened articulatory movements, particularly on targets in word-initial and word-final positions. In particular, three sound changes appeared significantly more often in articulation test responses: lengthening of vowels and consonants, additions of sounds represented by synchronic ties, and use of off-glides in productions of word-final vowels and consonants. In contrast, the allophones occurring in the conversational speech samples can be characterized as acting to assimilate and reduce sounds, evidenced by findings of increased use of the diacritic symbols for weakening, labialization, and dentalization of consonants and centralization of vowels.

Pass-Fail Analysis

A series of analyses was conducted for a perspective on the types of research and clinical decisions made from findings in each of the two speech sampling modes. The model for these analyses was the clinical paradigm of a pass-fail analysis, wherein each subject's status on each of the consonant sounds is dichotomized as a pass or a fail, that is, as a sound currently mastered or not yet mastered. As there is no standard or "true" criterion against which to compare the sensitivity and specificity of each of the current sampling modes, the analyses were constructed to yield concordance rates between modes.

Procedures. Percentage correct summaries for each subject in each sampling mode were assembled to yield totals for each of the consonant sounds in word-initial and word-final singletons and clusters. Articulation accuracy in word-medial or intervocalic singletons or clusters was excluded from these analyses due to low frequencies of occurrence in the articulation test stimuli. As in the preceding analyses, all eligible occurrences of sounds in word-initial and word-final positions on the articulation test stimuli were included.

For each consonant sound in each of the two word positions, matrices were constructed to categorize the sound as 100% correct in one or both modes, 0–99% correct in one or both modes, or having missing data in one or both modes. The low absolute frequency of occurrence of some sounds in each word position required the use of the 100% correct criteria for "pass"; matrices with missing data in at least one cell were excluded from the analyses. The individual data from each of these matrices were summed to yield for each of the 24 sounds, the percentage of children whose data indicated (a) a pass on both sampling modes, (b) a fail on both modes, (c) a pass on articulation test, but fail on continuous speech, and (d) a fail on articulation test, but pass on continuous speech.

Results. As shown in Table 4, results for word-initial singletons and clusters and word-final singletons and clusters were similar. Averaging findings from the two word

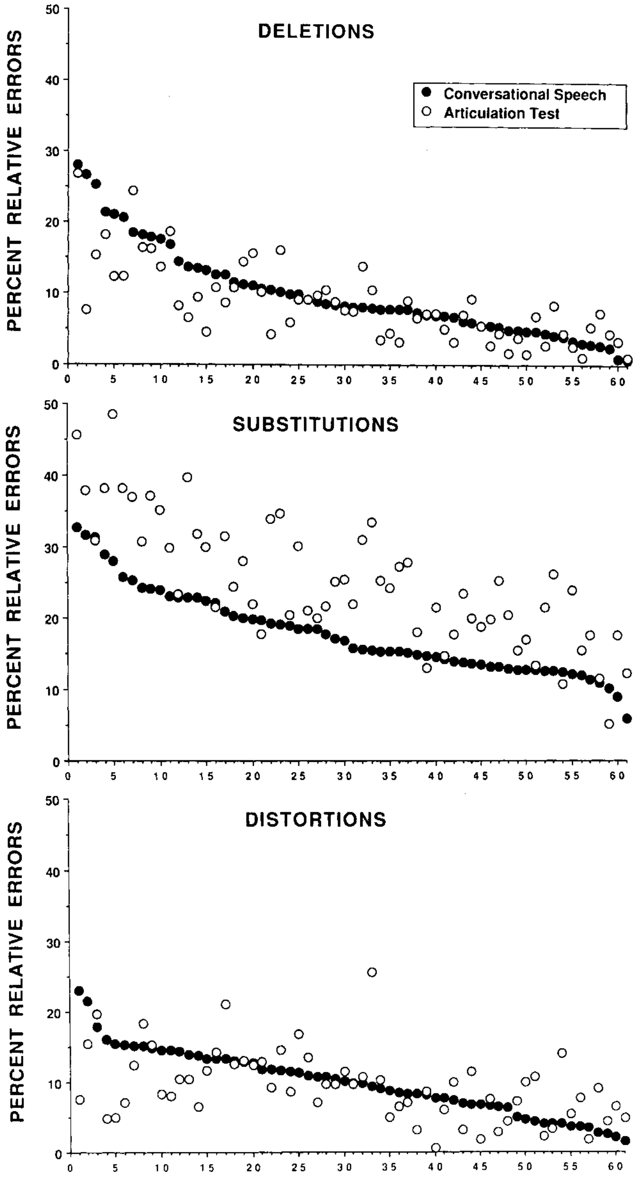

FIGURE 8. Relative error type analyses of consonant errors in conversational speech and articulation testing for individual subjects.

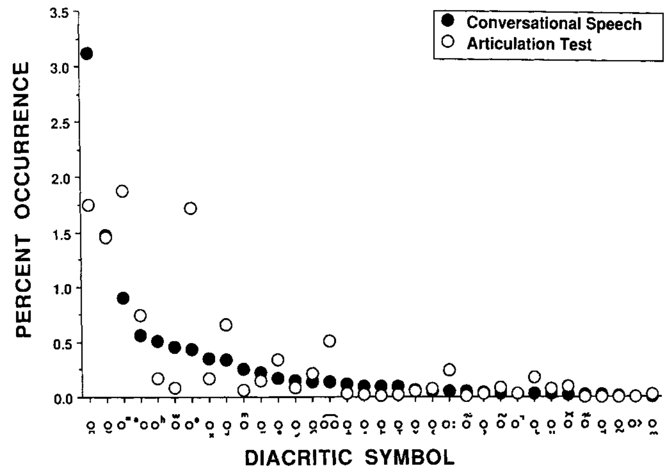

FIGURE 9. Subphonemic analysis of diacritic usage in conversational speech and articulation testing. The diacritic symbols (described in Shriberg & Kent, 1982) are arranged from right to left from most to least frequent in conversational speech.

positions, approximately 71% (17) of the consonant sounds received similar clinical decisions in the two sampling modes, approximately 22% (5) of the consonant sounds passed the articulation test, but failed on the continuous speech sample, and approximately 7% (2) of the consonant sounds failed the articulation test, but passed on the continuous speech sample. Thus, on average, dichotomous categorization based on the two sampling modes was similar for approximately two thirds of the 24 consonant sounds. Based on the standard deviations in Table 4, outcome discrepancies between the two sampling modes for two thirds of the 61 children occurred on as few as 13 sounds to as many as 20 sounds.

Figure 10 is a display of the sound-level information on the data presented in Table 4, with the top and lower panels including data for word-initial and word-final positions, respectively. For each consonant sound, the four bars indicate the percentage of subjects who passed both sampling modes, failed both modes, passed articulation testing but failed in continuous speech, and failed the articulation test but passed in continuous speech. Percentages for each sound were derived from the total number of children for whom sampling mode comparisons were possible. Consistent with the group and individual analyses, the data in Figure 10 indicate that the unconditional frequencies at which sounds are articulated correctly does not completely predict concordance rates (Kearns & Simmons, 1988). That is, concordance rates are not highest for all of the "easy" and "hard" sounds to articulate and lowest for sounds of intermediate difficulty. Rather, as divided into the four outcome categories, concordance status is marked by extensive interactions among sounds and word positions. The data in Figure 10 allow for a sound-by-sound inspection of these interactions. Review of the transcripts indicates that these subject-level concordance rates for each sound would be even more discrepant if broken out by consonant singletons and consonant clusters.

TABLE 4. Pass-fail analysis for each of the 24 consonants articulated by 61 children in articulation testing and conversational speech.

Pass-fail status	Position of singleton or cluster in word	Percentage of sounds *M*	*SD*
Pass or fail both sampling modes	Initial	72.5	13.4
	Final	69.1	11.8
Pass articulation test; fail conversational speech	Initial	21.0	11.1
	Final	23.4	12.4
Fail articulation test; pass conversational speech	Initial	6.7	6.4
	Final	7.3	6.8

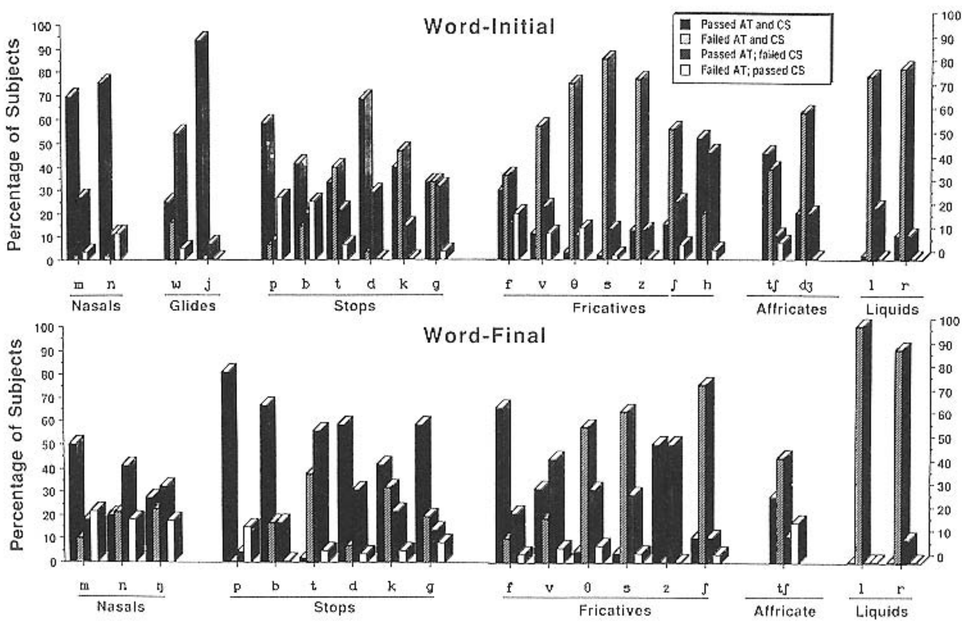

FIGURE 10. Pass-fail analyses of consonant singletons and clusters in articulation testing (AT) and conversational speech (CS). The upper panel includes the available data for word-initial position; the lower panel includes the available data for word-final position.

Individual Difference Analysis

A final series of analyses was designed to explore whether concordance between modes was associated with several individual difference variables that were available in the database.

Procedures. The data from Table 4 and Figure 10 were used to assign each of the 61 children to one of five categories for each word position. Skew and kurtosis values and associated plots supported a parametric approach to categorization, wherein children could be assigned to four groups based on their location in the distribution of scores relative to group means and standard deviations. Children whose percentage of inter-mode concordance for consonant sounds was above one standard deviation from the group were designated group A, children who scored within a standard deviation of the mean on each of the three agreement figures were designated group B, children who scored above one standard deviation from the group mean on percentage of sounds passed on the articulation test and failed in continuous speech were designated group C, children who scored above one standard deviation from the group mean on percentage of sounds failed on the articulation test and passed in continuous speech were designated group D, and children who scored above one standard deviation from the mean on both articulation pass/continuous speech fail and articulation test fail/continuous speech pass were designated group E. Group E was discarded because only one child in each word position met the description. Each of the other four groups included 7 to 34 children, with approximately half of the children meeting criteria for group B for both word-initial and word-final analyses.

With membership in one of the four groups used as the independent variable, Kruskal-Wallis One-Way Analyses of Variance were performed on each of six dependent variables: age, average words per utterance, percentage of vowels correct, percentage of singleton consonants correct, percentage of consonant clusters correct, and intelligibility. A total of 12 analyses were performed, with the data based on word-initial and word-final positions intended as cross-validational. The question was whether children's status on any of these demographic, language, or speech variables was associated with the relative number of sounds on which the two sampling modes yielded different pass/fail outcomes in each of the two word positions.

Results. Table 5 is a summary of the results of the individual differences analyses. No clear trends are apparent across the four groups. The only difference among the 12 comparisons to approach statistical significance, using a conservative α level adjusted for multiple tests, was for the word-initial data on Average Words Per Utterance ($H = 10.45$; $p < .02$), a language metric that correlates in the high .90s with mean length of utterance (Shriberg, 1986). The fact that the median data for word-final position are opposite in direction to the word-initial data for groups C and D—the groups with the largest median differences—suggests that this trend may not be reliable. Essentially, these final analyses indicated that none of the individual differences assessed was strongly associated with concordance rates for performance in the two sampling modes.

Discussion

Several methodological considerations should be kept in mind in review of these findings. First, the speech analyses were based solely on perceptual methods, with associated consequences for issues of validity and reliability (Shriberg and Lof, in press). Compared to data obtained by acoustic, kinematic, or physiological methods, each of which also has

TABLE 5. Individual difference analyses results based on the four-category, pass-fail data[a]. See text for an explanation of the four categories.

Category	Brief description	Age (months) Initial	Age (months) Final	Average words per utterance[b] Initial	Average words per utterance[b] Final	Percent correct vowels Initial	Percent correct vowels Final	Percent correct consonant singletons Initial	Percent correct consonant singletons Final	Percent correct consonant clusters Initial	Percent correct consonant clusters Final	Intelligibility (%) Initial	Intelligibility (%) Final
A	Above one standard deviation agreement between modes	53.0	52.0	4.1	3.0	92.4	92.2	64.0	58.1	50.0	40.1	92.5	82.0
B	Within one standard deviation agreement between modes	49.5	48.0	3.7	3.5	92.0	92.4	66.1	66.7	52.1	50.5	92.2	91.3
C	Above one standard deviation for sounds passed on articulation test and failed in conversational speech	55.0	55.0	4.6	3.7	92.9	92.1	64.6	64.8	45.5	51.3	90.0	91.8
D	Above one standard deviation for sounds failed on articulation test and passed in conversational speech	49.5	54.0	3.2	4.3	90.6	91.1	67.2	73.6	48.1	55.6	79.2	92.7

Note. [a] All table entries are medians; speech entries are taken from the conversational speech data. [b] Word-initial: $p < .05$ (Kruskal-Wallis H value adjusted for ties).

associated validity and reliability constraints, the present data rely solely on the auditory-perceptual vigilance and consensus decisions of persons well trained in phonetic transcription of disordered speech. Second, as reviewed previously, the continuous speech and citation form data were provided by different research transcription teams. Although detailed examination of transcription reliability for these consensus transcription teams suggests excellent interteam agreement at the level of broad phonetic transcription, transcription differences are a potential source of variance. Third, the findings are based on comparisons of continuous speech data to only one of the dozens of standardized articulation tests, using whole-word transcription rather than transcriptions of only the targeted sounds. Lastly, there were asymmetries in the token frequencies of sounds, with more tokens for most, but not all, sounds available from the conversational speech samples.

Research Considerations

The clear findings of this study are that the two sampling modes yield significant differences in the speech profiles of speech-delayed children. As in prior literature, the group-level and individual graphs and statistical effects do not all indicate that articulation testing yields higher performance scores. Rather, sampling mode differences in both directions were obtained at the linguistic levels of sound class, manner feature, phonological process, phoneme, error type, word position, and allophone. The complex pattern of speech sampling findings defies a consolidated explanation; that is, each finding cannot be neatly attributed to collateral effects associated with either perceptual (transcriber) or production (speaker) processes. Rather, an eventual account of the influence of sampling mode on research and clinical speech data will likely require a multifactorial framework. Therefore, the following few comments only extend the research implications for sampling, transcriber, and speaker variables discussed previously.

The specific structural and content differences between and within the two speech sampling modes assessed in this study are difficult to capture in one framework. Put most generally, in contrast to the monosyllabic and multisyllabic nouns that comprised the articulation test stimuli, the simple canonical structures in the recurrent function words occurring in the conversational speech included proportionally more of the earlier developing consonant sounds. Accordingly, a significant portion of the variance in the findings displayed in Figures 5 and 6 is assumed to reflect asymmetries in the lexical contexts for earlier and later occurring consonants, with the trend for the earlier developmental sounds to be more correct in continuous speech due to the lexical stimuli in which they are tested. The present database is not appropriate for a well-balanced inspection of the validity of this hypothesis. Controlled studies would require adequate tokens of comparable stimuli in each mode.

Potential variance associated with transcriber tasks focuses on the relative perceptual salience of articulatory targets in relation to the maintenance of stable response definitions for correct and incorrect articulation. The cognitive-perceptual demands of sentential transcription compared to word-level phonetic transcription tasks have not been explored. Stress-related sound changes in continuous speech production, such as reduced loudness and segment durations, may be associated with more liberal response definitions. In contrast, the clarity of word boundaries in single-word articulation tests may require subtle but systematically more articulate production to be considered correct. Moreover, in the richer contrasts provided in continuous speech, a child's error pattern on stressed compared to unstressed contexts could influence transcribers' perceptual criteria for both contexts. Thus, the relatively homogeneous contexts of repeatedly stressed nouns in citation forms may provide the perceptual context for stable response definitions, compared to those engendered by the constantly changing stress patterns of continuous speech. Such potential transcription processes are undocumented in the clinical and research speech pathology literature (see Shriberg & Lof, 1991, for extended discussion).

The speaker's contribution to the sampling mode differences also involve the individual and interactive contributions of cognitive-linguistic and pragmatic processes. The pattern of increased consonant deletions in continuous speech, which is consistent with the relatively higher rates of constituent omissions reported in the grammatical analyses of language-disordered children, may reflect a necessary shift in attention and speaking strategies across sampling contexts. In the articulation test format, emphasis is placed on lexical and phonological accuracy, that is, on retrieving the appropriate word and articulating it correctly. In contrast, in conversational samples, emphasis shifts to maintaining the accuracy and flow of ideas, often at the expense of less salient or important elements of form. In the present data, children produced developmentally earlier sounds better in spontaneous conversational speech and developmentally later sounds better in response to articulation test stimuli. Thus, when composed of familiar meaning, intentions, and forms, spontaneous conversational speech may provide the most optimal setting for production of well-established sounds or structures. In contrast, for less well-established sounds, the extensive linguistic, motor-speech, and pragmatic demands of that context may lead to decreased articulatory accuracy.

Clinical Considerations

Clinical implications of these data warrant brief comment, as most intervention programs include stages progressing from single-word training to naturalistic continuous speech. The present findings are consistent with the clinical observation that single-word training often is associated with exaggerated speech, including lengthening of both the target sound and the vowel nucleus, addition of on-glides and off-glides at word boundaries, affrication of initial and final obstruents, and other featural and segmental distortions. Such speaker behavior may, of course, at least partly mirror exaggerations in the stimuli provided for imitation by the examiner or speech-language pathologist. In contrast, fast or casual speech requires reduction, not addition, of features; the transitions from word to word, which turn final consonants into syllable-releasing segments, cannot be exaggerated without a disruption in prosody.

For children pressed to progress from single-word to continuous speech tasks, their options may be either to produce the newly learned forms with exaggerated beginnings and endings or delete the segments entirely. In the present study, transcribers reported (annotated on the transcript) a "choppiness" or discontinuity between words in the continuous speech of many of these speech-delayed children. These comments alluded to aspects of precision and timing that reflect the integration of segmental processing with suprasegmentals, including lexical, phrasal, and emphatic stress. Such observations are not unfamiliar to speech-language pathologists, wherein deletion of well-established consonants such as /t/, /d/, and /n/ and stopping of well-established sounds at word boundaries occurs as soon as a child moves to the segmental and suprasegmental demands of conversational speech. With approximately 80% of this clinical population having associated language production problems, and approximately 25% having associated prosody-voice involvement (Shriberg, 1991; Shriberg et al., 1986; Shriberg, Kwiatkowski, & Rasmussen, 1990), constraints in continuous speech could be associated with many levels of psycholinguistic and motor-speech processing. Conversational speech samples would appear to be the only source of *integrated* speech, language, and prosodic analyses needed to assess, plan intervention for, and monitor the progress of these children's individual phonological error patterns.

Conclusions

Reliable estimates of the interactive effects of sample, transcriber, and speaker variables in speech assessment will require carefully designed and controlled studies, moving well beyond the level of descriptive findings available from analyses of the present data and those cited throughout this report. Notwithstanding the lack of a comprehensive explanatory account, the current findings do suggest that attention to the potential effects of sampling mode is a crucial methodological need in child phonology research. As reviewed previously, single-word articulation tests and citation-form protocols continue to be used as the primary speech sampling mode for many contemporary research programs in developmental and nondevelopmental phonological disorders. Apparently, such methodological decisions are defended on utility and efficiency criteria and especially on the assumption of equivocal speech sampling findings in the archival literature. The present findings suggest that, in fact, statistically, clinically, and potentially theoretically significant speech performance differences are associated with sampling mode. If these subjects' speech had been sampled using only citation-form stimuli, some portion of the data and subsequent conclusions would be incomplete and possibly misleading relative to their performance in conversational speech samples obtained during the same assessment session.

Put most strongly, the present findings suggest that citation-form testing yields neither typical nor optimal measures of speech performance. Thus, as long as researchers and practitioners continue to base assessment results solely on different citation form tests, including the standardized articulation tests that continue to proliferate, sampling procedures will continue to be a major source of bias in phonological assessment (Butcher, 1990; Smit, 1990, 1991). Especially in view of emerging developments in nonlinear phonological theories, with their promise for integrated analyses of speech-language-prosody (Goldsmith, 1990; McGregor & Schwartz, 1991; Shriberg, 1990; Shriberg et al., 1990) and the utility of conversational speech samples for measures of intelligibility and severity of involvement (e.g., Garrett & Moran, in press), conversational speech sampling would seem to be the measurement procedure of choice. For continuous speech sampling of children with marked intelligibility problems, Kwiatkowski and Shriberg (1991) describe procedures that might be useful to augment examiner glosses with those provided by a caregiver.

Acknowledgments

We express our sincere thanks to a number of persons who provided competent research and editorial assistance at different stages of this work: Barri Babow, Maria Cavicchio, Patricia Engebose, Frederic Gruber, Rebecca Hinke, Sara Hoffman, Gregory Lof, Jane Loncke, Amparo Ortiz, Carmen Rasmussen, Dorothy Rorick, Catherine Trost-Steffen, and Carol Widder. We also acknowledge the continuing support of the Madison Metropolitan School District, Integrated Student Services, with special thanks to Beth Daggett and Susan Albert. This work was supported by grants from the United States Department of Education, G008400633 and the National Institute on Deafness and Other Communicative Disorders, DC00496.

References

Andrews, N., & Fey, M. E. (1986). Analysis of the speech of phonologically impaired children in two sampling conditions. *Language, Speech, and Hearing Services in Schools, 17,* 187–198.

Bankson, N. W., & Bernthal, J. E. (1982). A comparison of phonological processes identified through word and sentence imitation tasks of the PPA. *Language, Speech, and Hearing Services in Schools, 13,* 96–99.

Bernthal, J. E., & Bankson, N. W. (1988). *Articulation and phonological disorders* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Bernthal, J. E., Grossman, F. M., & Goll, A. E. A. (1989). Phonologically delayed children's responses to three types of pictured stimuli. *Journal of Childhood Communication Disorders, 12,* 137–143.

Blodgett, E., & Miller, V. (1986, November). *The effect of utterance length and imitation on disordered phonology.* Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Detroit.

Butcher, A. (1990). The uses and abuses of phonological assessment. *Child Language Teaching and Therapy, 6,* 262–276.

Campbell, T., & 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.

Chapman, R., & Ting, A. (1971). The effect of mode of elicitation in articulation testing. *Technical Report No. 154.* Wisconsin Research and Development, Center for Cognitive Learning, University of Wisconsin-Madison.

Compton, A. J., & Hutton, S. S. (1978). *Compton-Hutton Phonological Assessment.* San Francisco: Carousel House.

DuBois, E., & Bernthal, J. E. (1978). A comparison of three methods for obtaining articulatory responses. *Journal of Speech and Hearing Disorders, 43,* 295–305.

Dunn, C. (1982). Phonological process analysis: Contributions to assessing phonological disorders. *Communicative Disorders, 7,* 147–163.

Dunn, C., & Davis, B. (1983). Phonological process occurrence in phonologically disordered children. *Applied Psycholinguistics, 4,* 187–207.

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.

Elbert, M., Dinnsen, D. A., Swartzlander, P., & Chin, S. B. (1990). Generalization to conversational speech. *Journal of Speech and Hearing Disorders, 55,* 694–699.

Faircloth, M. A., & Faircloth, S. R. (1970). An analysis of the articulatory behavior of a speech-defective child in connected speech and in isolated-word responses. *Journal of Speech and Hearing Disorders, 35,* 51–61.

Fudala, J. B. (1974). *Arizona Articulation Proficiency Scale* (Rev. ed.). Los Angeles: Western Psychological Services.

Gallagher, R., & Shriner, T. (1975). Contextual variables related to inconsistent /s/ and /z/ production in the spontaneous speech of children. *Journal of Speech and Hearing Research, 18,* 623–633.

Garrett, K. K., & Moran, M. J. (in press). A comparison of phonological severity measures. *Language, Speech, and Hearing Services in Schools.*

Goldman, R., & Fristoe, M. (1969;1972). *Goldman-Fristoe Test of Articulation.* Circle Pines, MN: American Guidance Service.

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology.* Oxford, UK: Basil Blackwell, Ltd.

Harris, J., & Cottam, P. (1985). Phonetic features and phonological features in speech assessment. *British Journal of Disorders of Communication, 20,* 61–74.

Healy, T. J., & Madison, C. L. (1987). Articulation error migration: A comparison of single word and connected speech samples. *Journal of Communication Disorders, 20,* 129–136.

Henderson, F. M. (1938). Accuracy in testing articulation of speech sounds. *Journal of Educational Research, 31,* 348–356.

Hodson, B. (1980). *The assessment of phonological processes.* Danville, IL: The Interstate Press.

Hymes, D. H. (1971). Competence and performance in linguistic theory. In R. Huxley & E. Ingram (Eds.), *Language acquisition: Models and methods* (pp. 3–28). New York: Academic Press.

Johnson, J., Winney, B., & Pederson, O. (1980). Single word versus connected speech articulation testing. *Language, Speech, and Hearing Services in Schools, 11,* 175–179.

Jordan, E. P. (1960). Articulation test measures and listener ratings of articulation defectiveness. *Journal of Speech and Hearing Research, 3,* 303–319.

Kearns, K. P., & Simmons, N. N. (1988). Interobserver reliability and perceptual ratings: More than meets the ear. *Journal of Speech and Hearing Research, 31,* 131–135.

Kenney, K., Prather, E., Mooney, M., & Jeruzal, N. (1984). Comparisons among three articulation sampling procedures with preschool children. *Journal of Speech and Hearing Research, 27,* 226–231.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: The MIT Press.

Klein, H. B. (1984). Procedure for maximizing phonological information from single-word responses. *Language, Speech, and Hearing Services in Schools, 15,* 267–274.

Klein, H. B., & Spector, C. (1981, November). *The relationship between serial-position/stress level interaction and consonant errors.* Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Los Angeles.

Klein, H. B., & Spector, 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.

Kohn, S. E. (1988). Phonological production deficits in aphasia. In H. A. Whitaker (Ed.), *Phonological processes and brain mechanisms* (pp. 93–117). New York: Springer-Verlag.

Kwiatkowski, J., & Shriberg, L. D. (in press). Intelligibility assessment in developmental phonological disorders: Accuracy of caregiver gloss. *Journal of Speech and Hearing Research.*

Leopold, W. (1947). *Speech development of a bilingual child. A linguist's record. Vol. 2. Sound-learning in the first two years.* Evanston, IL: Northwestern University Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: The MIT Press.

Locke, J. (1983). *Phonological acquisition and change.* New York: Academic Press.

Lowe, R. (1986). *Assessment link between phonology and articulation.* Moline, IL: Lingui-Systems.

McGregor, K. K., & Schwartz, R. G. (in press). Converging evidence for underlying phonological representation. *Journal of Speech and Hearing Research.*

Menyuk, P. (1980). The role of context in misarticulations. In G. Yeni-Komishian, J. Kavanaugh, & C. Ferguson (Eds.), *Child phonology: Vol. 1. Production* (pp. 211–226). New York: Academic Press.

Miller, J. (1981). *Assessing language production in children.* Austin, TX: Pro-ed.

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *Journal of the Acoustical Society of America, 87,* 2705–2715.

Orr, E., Blodgett, E., & Miller, V. (1983, November). *Phonological analysis of spontaneous and imitated speech samples.* Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Cincinnati.

Paden, E., & Moss, S. (1985). Comparison of three phonological analysis procedures. *Language, Speech, and Hearing Services in Schools, 16,* 103–109.

Paynter, E., & Sims, F. (1979, November). *A comparison of two types of articulation assessment.* A paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Atlanta.

Pendergast, K., Dickey, S., Selmar, J., & Soder, A. (1969). *Photo Articulation Test* (2nd ed.). Danville, IL: The Interstate Publishers.

Schmitt, L. S., Howard, B. H., & Schmitt, J. F. (1983). Conversational speech sampling in the assessment of articulation proficiency. *Language, Speech, and Hearing Services in Schools, 14,* 210–214.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: The MIT Press.

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

Shriberg, L. D. (1990). Measurement validity. In L. B. Olswang, C. K. Thompson, S. F. Warren, & N. J. Minghetti (Eds.), *Treatment efficacy research in communication disorders* (pp. 63–78). Rockville, MD: American Speech-Language-Hearing Foundation.

Shriberg, L. D. (1991). Directions for research in development phonological disorders. In J. F. Miller (Ed.), *Research on child language disorders: A decade of progress* (pp. 267–276). Austin, TX: Pro-ed.

Shriberg, L. D., Hinke, R., & Trost-Steffen, C. (1987). A procedure to select and train persons for narrow phonetic transcription by consensus. *Clinical Linguistics and Phonetics, 1,* 171–189.

Shriberg, L. D., & Kent, R. D. (1982). *Clinical phonetics.* New York: Macmillan.

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

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonologic 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., Best, S., Hengst, J., & Terselic-Weber, B. (1986). Characteristics of children with speech delays of unknown origin. *Journal of Speech and Hearing Disorders, 51,* 140–160.

Shriberg, L. D., Kwiatkowski, J., & Hoffmann, K. (1984). A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Research, 27,* 456–465.

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

Shriberg, L. D., & Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics, 5,* 225–279.

Shriberg, L. D., & Olson, D. (1987). *PEPAGREE: Programs to compute transcriber agreement.* Madison, WI: Waisman Center on Mental Retardation and Human Development.

Siegel, R., Winitz, H., & Conkey, H. (1963). The influence of testing instruments on articulatory responses in children. *Journal of Speech and Hearing Disorders, 28,* 67–76.

Simmons, R., Blodgett, E., & Miller, V. (1983, November). *Assessment of phonological disorders in conversation and single word picture naming.* Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Cincinnati.

Smit, A. B. (1990, November). *Three kinds of variability (inconsistency) in children's speech.* Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Seattle.

Smit, A. B. (1991). *Children's mimicry of voiced and voiceless English stops: An investigation of relative ease of production.* Manuscript submitted for publication.

Snyder, L. (1984). Communicative competence in children with delayed language development. In R. Schiefelbusch (Ed.), *The acquisition of communicative competence* (pp. 423–478). Baltimore: University Park Press.

Stemberger, J. (1988). Between-word processes in child phonology. *Journal of Child Language, 15,* 39–61.

Stoel-Gammon, C. (1987). Phonological skills of 2-year-olds. *Language, Speech, and Hearing Services in Schools, 18,* 323–329.

Stoel-Gammon, C. (1988). Evaluation of phonological skills in preschool children. *Seminars in Speech and Language, 9,* 15–25.

Templin, M. C., & Darley, F. L. (1969). *The Templin-Darley Tests of Articulation* (2nd ed.). Iowa City: Bureau of Educational Research and Service.

Van Demark, D. R. (1964). Misarticulations and listener judgments of the speech of individuals with cleft palates. *Cleft Palate Journal, 1,* 232–245.

Vihman, M., & Greenlee, M. (1987). Individual differences in phonological development: Ages one and three years. *Journal of Speech and Hearing Research, 30,* 503–521.

Watson, M. M. (1989). Comparison of three methods for eliciting phonological processes. *Perceptual and Motor Skills, 69,* 771–778.

Weiner, F. (1979). *Phonological Process Analysis.* Baltimore: University Park Press.

Weiss, C. (1978). *The Weiss Comprehensive Articulation Test.* Boston: Teaching Resources.

Wren, C. T. (1985). Collecting language samples from children with syntax problems. *Language, Speech, and Hearing Services in Schools, 16,* 83–102.

Received January 31, 1991
Accepted July 29, 1991

Contact author: Lawrence D. Shriberg, PhD, Phonology Project, Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, WI 53705.