JSLHR, Volume 40, 723–740, August 1997

The Speech Disorders Classification System (SDCS): Extensions and Lifespan Reference Data

Lawrence D. Shriberg
Diane Austin
University of Wisconsin-Madison

Barbara A. Lewis
Case Western Reserve University
Cleveland, OH

Jane L. McSweeny
David L. Wilson
University of Wisconsin–Madison

A companion paper includes rationale for the use of 10 metrics of articulation competence in conversational speech (Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997). The present paper reports lifespan reference data for these measures using records from a total of 836 3- to 40+-year-old speakers with normal and disordered speech. The reference data are subdivided by diagnostic classification based on extensions to an instrument titled the Speech Disorders Classification System (SDCS; Shriberg, 1993). Appendices provide procedural information on the SDCS and statistical rationale for the reference data.

KEY WORDS: phonology, articulation, speech disorders, assessment, classification

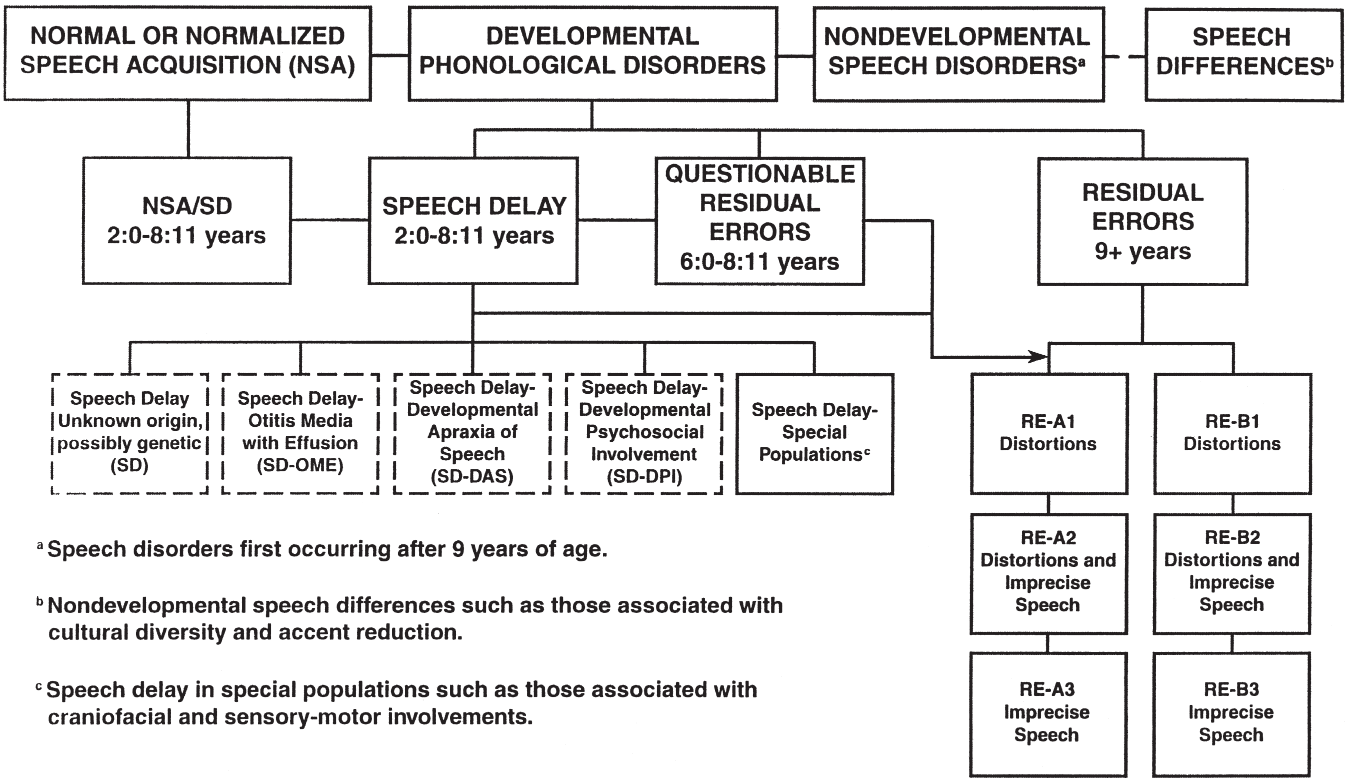
A previous paper provided rationale and validity data for a clinical classification instrument titled the *Speech Disorders Classification System* (SDCS; Shriberg, 1993). The present report provides rationale for extensions to the SDCS and lifespan reference data for the 10 measures of articulation competence described in a companion paper (Shriberg, Austin, et al., 1997). For the interested reader, Appendix A provides procedural detail for the SDCS program, and Appendix B provides rationale and procedural detail for statistical analysis of the lifespan reference data.

Extensions to the Speech Disorders Classification System (SDCS)

Figure 1 is a revised version of the SDCS described in Shriberg (1993, 1994), with extensions reflecting empirical findings in several subsequent reports (Lewis & Shriberg, 1994; Shriberg, Aram, & Kwiatkowski, 1997a, 1997b; Shriberg, Gruber, & Kwiatkowski, 1994; Shriberg & Kwiatkowski, 1994; Shriberg, Kwiatkowski, & Gruber, 1994). Discussions of the need for an etiologically based classification system, evaluative literature reviews, and precursors of the SDCS are presented elsewhere (cf. Shriberg, 1980, 1983, 1994, 1997; Shriberg & Austin, in press). The purpose of the SDCS is to classify a person's speech production status throughout the lifespan. Because the system was used to determine the speech status of persons in the reference data to be reported, it is useful to review theoretical and methodological assumptions underlying its original development and current extension. The following section provides

 1092-4388/97/4004-0723

Figure 1. Revised version of the Speech Disorders Classification System (SDCS). The solid lines depict descriptive subtypes; the dashed lines depict putative etiologic classifications.



an overview of those perspectives, and Appendix A provides a technical review of program steps. As shown in Figure 1, the SDCS system is organized by both *descriptive subtypes* and *etiologic subtypes*.

Descriptive Subtypes

The boxes in Figure 1 represent descriptive (solid lines) and etiologic (dashed lines) subtypes of speech disorders based on a speaker's current and prior speech production status as sampled in conversation (cf. Shriberg 1993, 1994). The upper leftmost box includes two types of speakers: those who have had normal acquisition of speech and those whose speech disorders have normalized (*normal or normalized speech acquisition:* NSA). The two upper rightmost boxes include adult and child speakers whose speech patterns are classified as *nondevelopmental speech disorders* and *speech differences*. Nondevelopmental speech disorders occur after the developmental period for speech acquisition, nominally after age 9. Speech differences include all speech and prosody-voice patterns that speakers may seek help in changing, such as patterns associated with multicultural diversity and the acquisition of English as a second language. Competent speech clinicians make clear classification distinctions between *differences* and *disorders*, although procedures used to help persons modify speech differences are generally similar to procedures used to normalize speech disorders.

The middle box in the upper row of Figure 1, *developmental phonological disorders* (DPD), is divided into 3 descriptive types: speech delay (SD), questionable residual errors (QRE), and residual errors (RE).

Speech Delay (SD)

SD is the classification for children younger than 9 years who have persisting deletion and substitution errors on low intelligibility (≤75%) not observed in typically speaking children of the same chronological age. The four putative etiologic subtypes of SD shown in dashed lines are described elsewhere (Shriberg, 1994, 1997; Shriberg & Austin, in press).

Questionable Residual Errors (QRE)

QRE is the classification for children age 6 years to 8 years 11 months whose error patterns include only common clinical distortions (or substitutions that are difficult to discriminate from common clinical distortions, e.g., w/r is similar to derhotacized /r/; see list of QRE substitutions in Appendix A). Some children with

QRE could be children who normalized from SD or NSA/SD status, or, as indicated in Figure 1, they could have no such history. Such errors *may* normalize by age 9, or they could persist.

Residual Errors (RE)

RE is the classification for speakers 9 years old and older who continue to have speech errors. RE includes two subtypes, RE-A and RE-B (Shriberg, 1993, 1994). Children and adults classified as RE-A have histories of SD, whereas speakers classified as RE-B do not have such histories. The three numbered subtypes classify the error type patterns: RE-A1 or RE-B1 is for residual common distortion errors only (as listed in Shriberg, 1993, Appendix), RE-A2 or RE-B2 is for residual common distortions and imprecise speech (omissions and substitutions; cf. Shriberg, 1993), and RE-A3 or RE-B3 is for imprecise speech. The RE-2 and RE-3 forms of residual errors may have very low prevalence in adult populations. The utility of these RE subtypes for descriptive-explanatory research is illustrated in Shriberg, Aram, et al. (1997a, 1997b), in which SDCS classifications are used to classify the speech status of children ranging in age from approximately 3 to 15 years.

New Classification Category for Speech as a Continuous Trait

In addition to the SDCS descriptive subtypes of NSA, SD, QRE, and RE (with its six subtypes), an additional classification was found to be needed in studies using the SDCS approach since 1993 (Shriberg). Although clinical concerns dictate that children be classified as normal (NSA) or speech delayed (SD) during the developmental period, this dichotomy is imposed on an underlying continuous verbal trait. For both research and clinical purposes, it is appropriate to provide a classification category for error patterns intermediate between NSA and SD. As shown in Figure 1, a descriptive subtype termed NSA/SD has been added to those described in Shriberg, 1993. The slash between the two stems indicates that NSA/SD truly falls between NSA and SD. As discussed later, alternative perspectives on children classified as NSA/SD depend on the particular purpose for which SDCS classification is being used.

Special Populations

The remaining SDCS classification in Figure 1 (indicated by solid lines) represents a placeholder for all other forms of speech delay found in *special populations* (a term used in Bernthal & Bankson, 1994). Etiologic subtypes in this category of special populations include speech delay associated with frank deficits in mechanism (e.g., tracheostomy), cognitive-linguistic (e.g., cognitive disability), and psychosocial (e.g., elective mutism) processes. Children in these special populations require special types of services and have specific types of speech involvements that differ from other forms of speech delay. However, because their deficits occur within the developmental period for speech acquisition, their profiles of speech delay may also be similar in many ways to the speech involvements of children in each of the other proposed etiologic subgroups. Thus, for both theoretical and applied educational purposes, it is useful to include this aggregate classification category within the SDCS system.

SDCS Classifications

Six Conventions Subserving Validity and Reliability Needs

Classification of a transcript into one of the descriptive subtypes above is performed by SDCS software using rules tables developed from review of normative data on speech-sound acquisition (cf. Shriberg, 1993, Appendix). Appendix A includes a graphic illustrating the decision rules used by the SDCS to classify a conversational speech sample as modified by the conventions described below. Although the etiologic subtypes are central in associated research, the SDCS software currently makes only *descriptive* subtype classifications. That is, to date the SDCS does not have the diagnostic markers to discriminate among the four etiologic subtypes of speech delay shown in dashed lines in Figure 1 (cf. Shriberg & Austin, in press). To understand the reference data to follow, it is important to describe six validity and reliability considerations used in the classification process. For ease of understanding these technical details, key concepts are italicized.

First, although the SDCS tabulates all speech errors, only age-inappropriate errors that occur in at least *two different words* are used for classification decisions. Moreover, because there are few developmental data on word-medial consonant development, errors in medial position are excluded from the classification decision process, as are sound changes termed casual forms (e.g., "I donno").

Second, to ensure the validity of NSA, NSA/SD, and QRE classifications, the *Intelligibility Index* (II) and the *number of usable words* in the sample are examined. Because the SDCS uses conservative criteria in making classification assignments (i.e., the default is NSA), a sample that is too brief or too unintelligible may not contain enough information for the program to assign an appropriate disorder classification. If the II is 75% or lower and the sample receives a preliminary classification of NSA, NSA/SD, or QRE, the sample is reclassified as SD! (with the ! indicating that the classification was made using intelligibility criteria). Similarly, if the

JSLHR, Volume 40, 723–740, August 1997

II is between 76% and 80% (inclusive) and the original classification is NSA or QRE, the sample is reclassified as NSA/SD!. If there are fewer than 100 usable words in a sample, an asterisk (*) is appended to the classification to indicate that detailed examination of the sample or a longer conversational speech sample may be needed to make an accurate classification.

Third, if a speaker has age-inappropriate errors for only one sound, or has an incomplete inventory of word shapes, consonants, or vowels/diphthongs, the NSA classification is enclosed in *curly braces* (i.e., {NSA}). Moreover, if a speaker has made age-appropriate errors on one or more sounds, a "-" is added to the NSA stem (i.e., NSA- or {NSA-}).

Fourth, when young children (aged 2 years to 5 years 11 months) meet criteria for SD, but would not be classified as SD if they were one year younger, the SD classification is prefixed by *Questionable* (i.e., QSD).

Fifth, when an older (9+ years) speaker's error pattern is between NSA and one of the RE types, the classification is enclosed in *square brackets* to indicate that it is a marginal classification (e.g., [RE-A1]). For younger children, marginal classifications are subsumed by either {NSA} or NSA/SD.

Sixth, when a speaker's error pattern includes uncommon clinical distortions in over 20% of words (see list of uncommon distortions in Shriberg, 1993), the *suffix* "+" is added to the classification stem (e.g., SD+, RE-A1+). A square bracket around the "+" (e.g., SD[+]) indicates that only 10% to 20% of the words in the sample contain these distortions.

Thus, in combination, these six conventions lead to several variants of each main classification type (e.g., {NSA-}*, QSD[+], NSA–/SD! [RE-A2][+]). For particular clinical and research applications, such classification detail is ignored or retained, depending on validity, reliability, and efficiency needs. For example, the sample reference data to follow does not differentiate among the prevalence percentages for bracketed versus unbracketed classifications.

Lifespan Reference Data

Articulation Competence Measures

A companion paper (Shriberg, Austin, et al., 1997) describes 10 measures of articulation competence that can be calculated from a conversational speech sample. Rationale is provided for the following alternatives to the Percentage of Consonants Correct (PCC): Percentage of Consonants Correct–Adjusted (PCC-A), Percentage of Consonants Correct–Revised (PCC-R), Percentage of Consonants in the Inventory (PCI), Articulation Competence Index (ACI), Percentage of Vowels/Diphthongs Correct (PVC), Percentage of Vowels/Diphthongs Correct–Revised (PVC-R), Percentage of Phonemes Correct (PPC), Percentage of Phonemes Correct–Revised (PPC-R), and the Intelligibility Index (II). For each of the consonant measures (PCC, PCC-A, PCC-R, ACI, PCI), subscale data are also provided by developmental sound class (Early-8, Middle-8, Late-8). Transcription reliability data and standard error of measurement data are provided for all measures and submeasures (including reliability of SDCS classification) with the exception of the Intelligibility Index.

A brief review of the characteristics of some of these measures is important to the discussion of the reference data. For three "original" measures of consonants (PCC), vowels (PVC), or phonemes (PPC) "revised" measures have been derived (i.e., PCC-R, PVC-R, PPC-R). For consonants, there is also an "adjusted" measure (PCC-A). The original, revised, and adjusted measures differ in the way they score speech-sound distortions. The original measures score all omissions, substitutions, and distortions as errors, whereas the revised measures score only omissions and substitutions as errors. The PCC-A makes a further distinction, dividing distortion errors into common clinical distortions and uncommon clinical distortions, and scoring only omissions, substitutions, and uncommon clinical distortions as errors.

Database

Since approximately 1980, essentially similar conversational speech protocols and data reduction procedures have been used to obtain, transcribe, and format speech samples from children and adults for computerized analyses (cf. Shriberg, 1986, 1993; Shriberg & Kent, 1982, 1995; Shriberg & Kwiatkowski, 1980, 1982, 1983, 1985; Shriberg, Kwiatkowski, & Hoffmann, 1984). In addition to cross-sectional and follow-up studies conducted in Madison and several other cities in Wisconsin, conversational speech samples from children and adults have been collected in collaborative research in several states. With few exceptions, the demographic characteristics of speakers from these sites are middle-class, monolingual English-speaking children and adults speaking General American dialect. From an initial database of 1,386 conversational speech samples, reference samples for each of the 10 speech measures to be reported were assembled in three steps.

First, the database was searched to eliminate conversational samples from children and adults with known developmental disability, cognitive disability, craniofacial anomaly, and/or sensory-motor problems. Additionally, if more than one conversational speech sample was available from a speaker (e.g., from a treatment or longitudinal study), the sample selected for inclusion was either the earliest, the most linguistically representative, or the sample that best met cell size needs for each subgroup in the reference data.

Second, for each age from 3 years to 8 years, conversational samples from four subgroups of children were assembled based on their SDCS classification: NSA, NSA/SD, SD, and QRE. (Only 6- to 8-year-old children are eligible for classification as QRE.) All bracketed SDCS classifications and "+" and "–" suffixes were disregarded, and QSD and SD classifications were combined. Four classification groups of older children and adults were also assembled: NSA, RE-1, RE-2, and RE-3. SDCS classifications were combined across brackets and suffixes, as were RE-A (history of speech delay) and RE-B (no history of speech delay) classifications.

Third, to maximize the validity of the NSA, NSA/SD, and QRE subgroup data, conversational samples for children with these SDCS classifications were excluded if there were fewer than 100 usable words in the sample. Because the SDCS program uses conservative criteria in making classification assignments (the default is NSA), a sample that was too brief might not have contained a sufficient number of tokens to be classified reliably as one of the SDCS disorder categories.

Table 1 summarizes cell sizes by age, gender, and SDCS group for the final set of reference data to be reported. The SDCS program classified a conversational speech sample from each of 836 speakers into one of the cells in Table 1. For 3- to 8-year-old children, cell sizes are reported for single-year age groups and for combined groups of 3- to 5-year-olds and 6- to 8-year-olds. For the 657 children aged 3 years to 8 years 11 months, cell sizes for the NSA group range from 7 (8-year-old boys and girls) to 59 (5-year-old boys). For the 179 older children and adults, cell sizes for the NSA group range from 5 (9- to 11-year-old girls) to 42 (18- to 39-year-old women).

Again, it is important to underscore that the reference data for these speakers are purposefully defined by SDCS classification categories, rather than by history of speech involvement, history of speech normalization, or suspected etiologic subtype. This procedure assures a standard definition for NSA and the disorder categories, rather than relying on clinical experience and historical data to define which speaker is normal and which is disordered. Thus, data for speakers coded NSA (normal or normalized speech acquisition) reflect performance from both control subjects in several studies (including adult family members of children with speech disorders) and speech-disordered subjects whose speech had normalized at the tested age. Speakers meeting SDCS criteria for SD were referred by speech-language pathologists, recruited as subjects, or identified as SD by the SDCS in the course of a study (e.g., a sibling of an SD child). As noted above, the many demographic

Table 1. Cell sizes for the reference data by age, SDCS classification, and gender.

	NSA			NSA/SD			SD			QRE			All		
Age	M	F	Total	M	F	Total	M	F	Total	M	F	Total	M	F	Total
Ages 3:0 –8:11															
3	22	30	52	11	4	15	34	12	46	–	–	–	67	46	113
4	28	19	47	22	13	35	51	24	75	–	–	–	101	56	157
5	59	50	109	20	11	31	39	17	56	–	–	–	118	78	196
3-5:11	109	99	208	53	28	81	124	53	177	–	–	–	286	180	466
6	45	31	76	15	7	22	15	3	18	11	3	14	86	44	130
7	9	14	23	3	1	4	5	1	6	3	0	3	20	16	36
8	7	7	14	2	1	3	1	0	1	6	1	7	16	9	25
6–8:11	61	52	113	20	9	29	21	4	25	20	4	24	122	69	191
Total	170	151	321	73	37	110	145	57	202	20	4	24	408	249	657
	NSA			RE-1			RE-2			RE-3			All		
Age	M	F	Total	M	F	Total	M	F	Total	M	F	Total	M	F	Total
Ages 9:0 –40+															
9–11:11	9	5	14	8	11	19	3	5	8	1	0	1	21	21	42
12–17:11	15	10	25	8	4	12	4	1	5	0	0	0	27	15	42
18–39:11	25	42	67	3	2	5	0	1	1	0	0	0	28	45	73
40+	9	9	18	1	2	3	1	0	1	0	0	0	11	11	22
Total	58	66	124	20	19	39	8	7	15	1	0	1	87	92	179

constraints relative to standard test development and epidemiological considerations require that the information assembled from these database samples be viewed as *reference data*, rather than as *normative* data. Thus, they provide only guidelines for use of the 10 measures. We caution readers against unquestioned use of these reference data for clinical decision making or for control data in research studies.

Statistical Analyses

A series of statistical analyses were undertaken to determine how best to summarize and present reference data for the 10 speech measures, including rationale for calculating *z* scores and the most appropriate standard score. The primary questions were whether the raw scores for all measures met distributional criteria for parametric descriptive statistics and/or whether transformations were needed to make the distributions more normally distributed. Step-wise rationale for this and other questions and all procedural details are summarized in Appendix B.

The statistical analyses summarized in Appendix B supported the use of parametric statistics (mean, standard deviation) when describing untransformed raw scores on all measures and the use of *z* scores or standard scores when comparing the performance of speakers of different ages and genders. They also supported the use of group-wise standard deviations termed *derived* standard deviations when calculating *z* scores and standard scores. The interested reader is referred to Appendix B for rationale underlying statistical options and details of these analyses.

Sample Reference Data for 3- to 8-Year-Old Children

Table 2 illustrates the format and content of the reference data available for 3- to 8-year-old children. The data in Table 2 are the reference data for 3-year-old children. A complete set of reference data is available for viewing or downloading from the Phonology Project Web site, and a hard copy is available by mail from the first author. The complete set of reference data includes descriptive statistics similar to those in Table 2 for children at each age from 3 to 8 years and for children/adults grouped into the age ranges of 9–11, 12–17, 18–39, and 40+. As described in Appendix B, most older speakers approach 100% on the 10 speech measures, making *z* scores and standard scores meaningless. However, for comparison with persons with nondevelopmental speech problems, the reference data for older speakers include the minimum score for each measure in each subgroup (cf. Appendix B).

Substantive comment on the reference data itself is beyond the scope of the present report. However, four findings in the reference data support the content, criterion, and construct validity of the SDCS classifications described in Shriberg (1993) and in this report.

Validity Support for SDCS Classification: NSA

First, as shown in Figure 2, the reference data provide both construct and criterion validity to the SDCS program's classification of a speech sample as NSA. NSA children in most of the Age × Gender subgroups (see Table 1 for cell sizes) had mean scores above 90% for most of the 10 speech measures. As shown, means of 90% or better were obtained for 3- to 8-year-old children (a) on the four measures that score all or some clinical distortions as correct (PCC-A, PCC-R, PVC-R, and PPC-R); (b) on the PCI, PVC, and II (including the Early-8, Middle-8, and Late-8 PCI subscales not shown in Figure 2); and (c) for the Early-8 and Middle-8 subscales of the PCC, PCC-A, and PCC-R (not shown in Figure 2). Excluding the ACI (which yields lower mean percentages because of the formula used to weight its components), the lowest means for 3- to 5-year-old children classified as NSA were 79.4% for the PCC and 85.4% for the PPC for 3-year-old boys, with means for 6- to 8-year-olds above 90% for both of these measures.

Thus, the developmental criteria used by the SDCS to classify children as NSA (i.e., based on age norms for speech-sound mastery) are closely associated with the quantitative data reflected on the 10 speech measures. As 90% competence or better on a measure is generally taken to be "within the normal range," these findings support the use of NSA means and standard deviations to compute *z* scores and standard scores (see below).

Validity Support for SDCS Classification: RE

A second database finding supports the content and construct validity of the RE classifications within the SDCS. Figure 3 is a summary of the percentages of each SDCS classification based on conversational speech samples from the 179 9- to 40+-year-old speakers. As shown, approximately 67% of the samples from male speakers and 72% of samples from the female speakers were classified as NSA. Of the remaining approximately 33% of the samples for males, 23% were RE-1, 9.2% were RE-2, and 1.1% were RE-3; for females, 20.7% were RE-1, 7.6% were RE-2, and 0% were RE-3.

Once again, demographic and sampling constraints prohibit viewing these data as estimates of the prevalence of each subtype of RE in the general population.

Table 2. Sample reference data: Ages 3:0 –3:11.

		Male											Female											Both					
		NSA			NSA/SD				SD				NSA			NSA/SD				SD				NSA		NSA/SD		SD	
Metric[a]	DC[b]	*M*[c]	*SD1*	*SD2*	*M*	*SD1*	*Z*	*S*	*M*	*SD1*	*Z*	*S*	*M*	*SD1*	*SD2*	*M*	*SD1*	*Z*	*S*	*M*	*SD1*	*Z*	*S*	*M*	*SD1*	*M*	*SD1*	*M*	*SD1*
PCC	E	97.5	2.4	2.3	93.0	3.2	−1.9	90.3	85.1	11.6	−5.4	73.1	97.9	1.9	1.5	95.2	5.1	−1.8	90.8	86.9	11.4	−7.4	63.2	97.7	2.1	93.6	3.7	85.6	11.5
	M	93.6	6.2	5.1	80.8	7.4	−2.5	87.5	63.4	14.7	−5.9	70.4	94.6	4.7	4.2	76.6	15.8	−4.3	78.4	62.2	14.3	−7.7	61.3	94.2	5.3	79.7	9.8	63.1	14.4
	L	41.6	13.5	20.7	24.2	15.9	−0.8	95.8	19.6	14.4	−1.1	94.7	49.0	17.4	20.5	34.5	14.1	−0.7	96.4	12.8	13.7	−1.8	91.2	45.9	16.1	26.9	15.6	17.9	14.4
	T	79.4	6.8	7.9	70.3	7.1	−1.2	94.2	61.9	11.6	−2.2	88.9	80.9	7.1	7.7	73.7	4.9	−1.0	95.2	60.7	9.6	−2.6	86.8	80.3	6.9	71.2	6.6	61.6	11.0
PCC–A	E	97.5	2.4	2.3	93.0	3.2	−1.9	90.3	85.1	11.6	−5.4	73.1	97.9	1.9	1.5	95.2	5.1	−1.8	90.8	86.9	11.4	−7.4	63.2	97.7	2.1	93.6	3.7	85.6	11.5
	M	94.3	5.2	4.9	81.3	7.4	−2.7	86.7	63.6	14.7	−6.3	68.6	95.4	4.3	4.1	76.6	15.8	−4.6	76.9	62.2	14.3	−8.1	59.5	94.9	4.7	80.0	9.9	63.2	14.4
	L	78.2	8.7	11.2	54.5	18.5	−2.1	89.4	33.4	17.8	−4.0	80.0	83.8	11.1	11.1	45.2	6.0	−3.5	82.6	30.6	17.8	−4.8	76.1	81.4	10.4	52.0	16.4	32.7	17.6
	T	91.0	3.9	4.4	79.4	5.2	−2.6	86.8	65.9	11.5	−5.7	71.5	92.6	4.7	4.3	76.5	4.9	−3.7	81.4	65.5	11.6	−6.3	68.6	91.9	4.4	78.6	5.1	65.8	11.4
PCC-R	E	98.3	1.5	1.8	94.3	3.1	−2.2	88.8	87.4	9.9	−6.0	69.9	98.3	1.6	1.3	95.6	4.7	−2.1	89.6	89.1	10.7	−7.1	64.6	98.3	1.5	94.6	3.5	87.9	10.0
	M	96.3	3.2	4.4	85.8	7.0	−2.4	88.1	67.2	13.6	−6.6	66.9	96.8	4.0	4.1	76.9	16.3	−4.8	75.8	66.7	15.4	−7.3	63.3	96.6	3.6	83.5	10.4	67.0	13.9
	L	81.6	9.0	11.4	57.9	18.7	−2.1	89.6	35.4	18.4	−4.0	79.8	86.9	11.4	11.5	45.4	6.5	−3.6	82.1	32.8	19.5	−4.7	76.6	84.6	10.7	54.6	17.0	34.7	18.5
	T	92.8	3.1	4.2	82.2	4.4	−2.5	87.3	68.5	10.4	−5.8	71.1	94.3	4.6	4.3	76.9	5.1	−4.0	79.8	68.4	11.8	−6.0	69.9	93.7	4.1	80.8	5.0	68.5	10.7
ACI	E	95.1	10.2	12.8	62.5	19.3	−2.5	87.3	56.2	18.5	−3.0	84.8	94.9	12.6	6.7	85.2	24.8	−1.4	92.9	66.6	24.4	−4.2	79.0	95.0	11.5	68.6	22.5	58.9	20.5
	M	80.2	20.0	22.3	52.8	9.1	−1.2	93.9	36.8	8.0	−1.9	90.3	85.5	19.1	21.5	41.0	12.6	−2.1	89.6	38.1	13.7	−2.2	88.9	83.3	19.5	49.7	11.1	37.1	9.6
	L	54.2	8.3	13.4	33.9	12.7	−1.5	92.5	19.7	10.8	−2.6	87.2	61.0	12.4	13.8	24.8	3.7	−2.6	87.0	18.1	11.4	−3.1	84.5	58.1	11.3	31.4	11.6	19.3	10.9
	T	70.9	8.1	12.8	53.9	7.5	−1.3	93.4	39.5	7.4	−2.4	87.8	74.5	10.4	13.3	43.0	5.8	−2.4	88.1	40.9	9.9	−2.5	87.4	73.0	9.6	51.0	8.5	39.9	8.0
PCI	E	100.0	0.0	0.4	100.0	0.0	0.0	100.0	97.1	6.2	−7.3	63.6	100.0	0.0	0.0	98.4	3.1	0.0	0.0	99.4	1.9	0.0	0.0	100.0	0.0	99.6	1.6	97.7	5.5
	M	96.5	5.1	6.0	86.1	11.7	−1.7	91.3	74.8	17.3	−3.6	81.9	97.0	5.6	4.2	86.3	11.2	−2.6	87.1	66.1	16.2	−7.4	63.2	96.8	5.4	86.2	11.1	72.6	17.3
	L	91.9	11.2	9.2	78.3	17.1	−1.5	92.6	58.2	24.0	−3.7	81.6	94.3	7.8	8.9	69.6	28.8	−2.8	86.0	55.5	30.3	−4.4	78.1	93.3	9.4	76.0	20.1	57.5	25.5
	T	96.4	3.4	3.9	88.8	6.7	−2.0	90.2	78.2	11.1	−4.7	76.7	97.2	3.6	3.4	85.4	13.5	−3.5	82.5	75.1	12.9	−6.5	67.5	96.9	3.5	87.9	8.6	77.4	11.5
PVC	94.4	2.7	2.4	91.0	5.4	−1.4	93.0	88.7	8.8	−2.3	88.3	94.3	4.0	3.1	95.5	2.9	0.4	102.0	87.8	6.2	−2.1	89.6	94.3	3.5	92.2	5.2	88.5	8.2	
PVC-R	97.9	1.7	1.8	95.3	4.5	−1.4	93.0	92.1	6.0	−3.2	84.0	98.0	2.7	1.8	97.0	2.0	−0.6	97.1	91.6	5.9	−3.6	82.1	98.0	2.3	95.8	4.0	92.0	5.9	
PPC	85.4	4.6	5.3	78.6	4.7	−1.3	93.5	72.8	10.0	−2.4	88.1	86.3	4.8	5.2	82.6	2.8	−0.7	96.5	72.0	7.4	−2.8	86.2	85.9	4.7	79.6	4.6	72.6	9.3	
PPC-R	94.9	2.2	2.9	87.5	3.6	−2.6	87.2	78.1	8.3	−5.8	71.1	95.8	3.2	2.9	85.0	2.9	−3.7	81.5	78.0	8.6	−6.1	69.3	95.4	2.8	86.9	3.5	78.1	8.3	
II	94.8	3.1	3.0	92.3	5.8	−0.8	95.8	79.5	12.1	−5.1	74.6	96.3	3.0	2.4	91.1	6.2	−2.2	89.0	85.9	10.8	−4.3	78.4	95.7	3.1	92.0	5.7	81.2	12.0	

[a]PCC: Percentage of Consonants Correct. PCC-A: Percentage of Consonants Correct–Adjusted. PCC-R: Percentage of Consonants Correct–Revised. ACI: Articulation Competence Index. PCI: Percentage of Consonants in the Inventory. PVC: Percentage of Vowels Correct. PVC-R: Percentage of Vowels Correct–Revised. PPC: Percentage of Phonemes Correct. PPC-R: Percentage of Phonemes Correct–Revised. II: Intelligibility Index.
[b]DC: Developmental Sound Class. E: Early-8. M: Middle-8. L: Late-8. T: Total for all consonants.
[c]*M*: mean. *SD1*: sample standard deviation. *SD2*: derived standard deviation. *Z*: *z* score. S: standard score.

Figure 2. Construct and criterion validity data for four classifications in the Speech Disorders Classification System (SDCS).

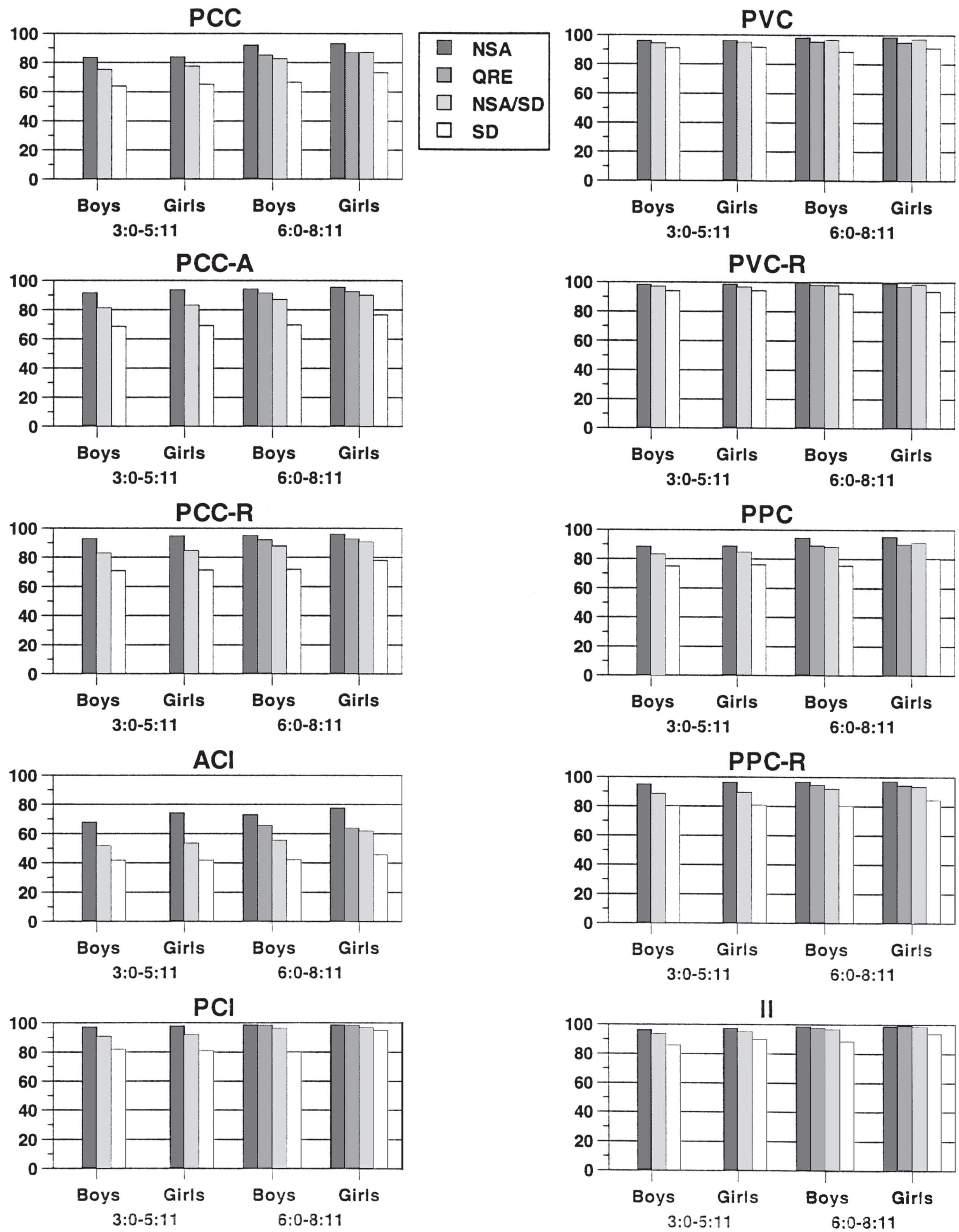
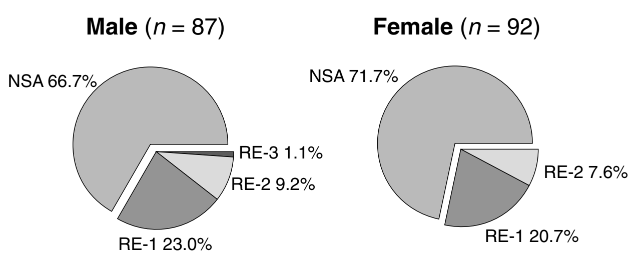


Figure 3. Content and construct validity data for the three Residual Errors (RE) subtypes in the Speech Disorders Classification System (SDCS).

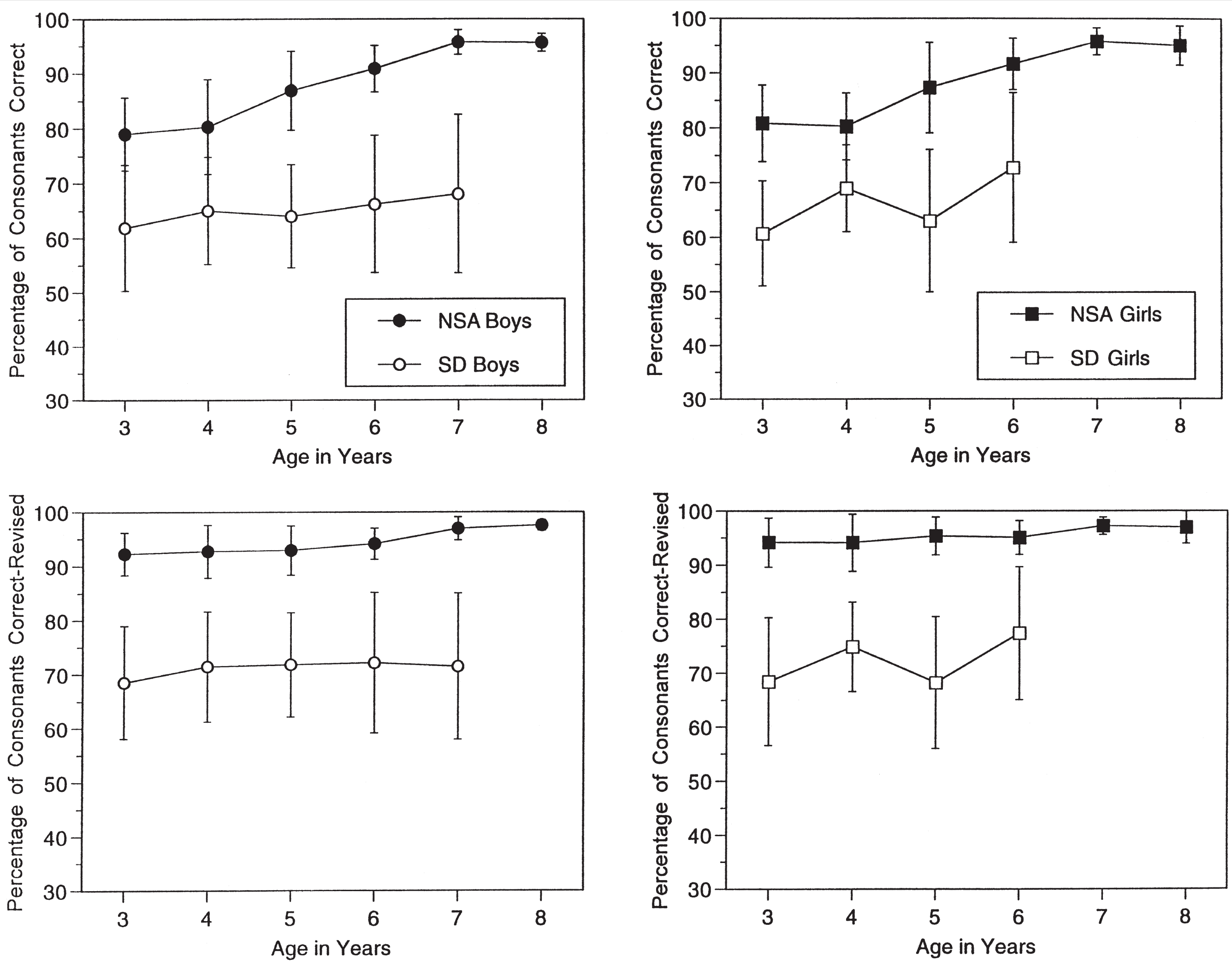


The relatively high percentages for some subtypes of RE reflect the emphasis in the database on speakers with speech disorders. However, because the three subtypes of RE were created to make the SDCS sensitive to speech disorders across the lifespan, these percentages do support content and construct validity of the SDCS. Specifically, at least one speaker in the database was classified as one of the three RE subtypes, RE-1, RE-2, and RE-3. Further information on the types of distortion errors for RE-1 and RE-2 classifications and information on error types by age and gender will be reviewed in the context of a forthcoming acoustic study of RE-A versus RE-B errors.

Validity Support for Use of the PCC-R Measure for Clinical Research

A third set of findings in the reference data supports the conceptual and psychometric features of the PCC-R—suggesting the PCC-R as the one best measure of articulation competence in 3- to 8-year-old children. Two sets of

Figure 4. Comparisons between Percentage of Consonants Correct (PCC) scores (upper panels) and Percentage of Consonants Correct–Revised (PCC-R) scores (lower panels) for 325 speech normal (NSA) and 200 speech-delayed (SD) boys and girls divided into Age x Gender subgroups.



data illustrate the properties of the PCC-R metric.

First, Figure 4 is a graph of the means and standard deviations for the PCC and PCC-R metrics for NSA and SD boys and girls from age 3 to 8. (Subgroups with only one individual are not displayed.) This figure is similar in concept to Figure 5 in Shriberg (1993), which contrasted scores on two measures (PCC and ACI) for children with normal and disordered speech. However, classification as normal versus disordered speech was based on clinical referral and history in Shriberg (1993), whereas the present classification is based on SDCS designation as NSA versus SD. The two left panels contrast PCC and PCC-R data for NSA and SD boys and the two right panels provide these comparisons for girls.

The panels in Figure 4 indicate that the PCC-R provides better separation of NSA scores from SD scores compared to the separation in the PCC scores. Note the overlapped standard deviations for the PCC scores of the 3- and 4-year-old boys and the 4-year-old girls (similar to findings in Shriberg, 1993), whereas there is no overlap for NSA versus SD groups in the PCC-R scores. The increased spread is associated with both more separation between means and generally smaller standard deviations for PCC-R scores. As discussed previously, at these young ages error types distinguish normally speaking children from speech-delayed children, with substitutions and deletions accounting for approximately 37% of the (generally fewer) errors for the NSA children, compared to 82% of the errors for the SD children. Additional support for differences between NSA and SD profiles at the level of distortions can be found in a comparison of Late-8 subscale profiles for children in each category on the three metrics, PCC, PCC-A, and PCC-R. For 3-year-old boys, for example, means for the NSA group for the Late-8 subscales of the PCC, PCC-A, and PCC-R are 41.6, 78.2, 81.6 respectively, with distortions accounting for 40 percentage points of the NSA group's lowered PCC score, and substitutions and deletions accounting for the remaining 19 percentage points. For the SD group, the corresponding Late-8 scores are 19.6, 33.4, and 35.4, with distortions accounting for only 15 percentage points of the lowered PCC scores, and substitutions and deletions accounting for approximately 65 percentage points.

Figure 4 also illustrates interesting age trends for the NSA boys and girls as reflected in the PCC compared to the PCC-R measures. Whereas the trend for mean PCC-R scores is essentially flat from 3 to 8 years for both boys and girls (ranging from 92.8% to 97.5%), trends during this 6-year developmental period are more sharply rising when measured with the PCC, starting at 79.4% and rising to 95.8%. The difference in measures is that the PCC reflects the allophonic mastery occurring during this period (i.e., includes distortions as errors), whereas the PCC-R more reflects phonemic mastery. It is predominantly the latter that distinguishes children with delayed speech from their typically developing age mates.

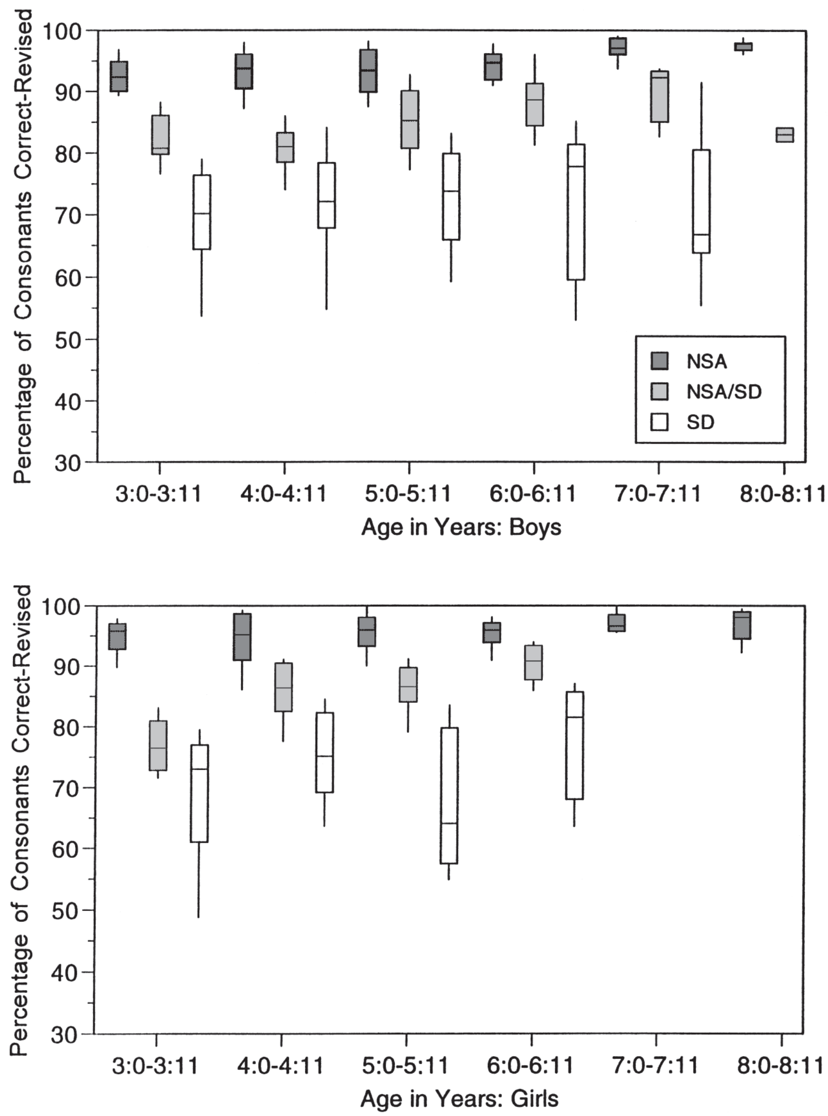
Figure 5 provides information on the construct validity of NSA/SD as an SDCS classification needed to measure speech as a continuous trait. The panels in Figure 5 include box plots of the PCC-R scores for the 3- to 8-year-old boys and girls, with separate box plots for the NSA, NSA/SD, and SD subgroups (subgroups with only one individual are not displayed). These data provide empirical support for the NSA/SD groups as intermediate to the NSA and SD groups in both qualitative status (SDCS classification) and quantitative score (PCC-R). Note that the interquartile ranges (25th–75th percentiles) for both boys and girls have minimal overlap among the three SDCS classification groups. The overlap of the NSA/SD and SD groups for 3-year-old girls is likely associated with the small cell size (4) for the NSA/SD group. Separation of the NSA from the SD groups is excellent, with overlap of only a few individual scores. Similar separations were observed on plots for PCC-A scores and PPC-R scores. For these two measures and the PCC-R, SD scores within one standard deviation of the mean for each Age × Gender subgroup do not overlap with NSA scores in the respective subgroups.

Validity Support for the Use of *z* Scores and Standard Scores Based on These Reference Data

The final series of analyses were undertaken to confirm conclusions from the statistical analyses described in Appendix B concerning the suitability of these reference data for *z* score and standard score transformations. The question was whether *z* scores for each of the three primary SDCS classifications (NSA, NSA/SD, SD) would have appropriate separation on the speech measures.

Z scores for all 3- to 8-year-old children in the NSA, NSA/SD, and SD groups were calculated, using the procedure recommended in Appendix B (i.e., using the Age × Gender NSA means and the derived NSA standard deviations). For the NSA/SD Age × Gender subgroups (excluding groups with only one or two members), the mean *z* scores for the PCC-A, PCC-R, and PPC-R ranged from –1.5 to –4.1; and the PVC-R and PCC *z* score means ranged from –.06 to –2.1. For the SD subgroups, mean *z* scores for the PCC-A, PCC-R, and PPC-R ranged from –4.4 to –9.2, with the PVC-R and PCC *z* score means ranging from –1.5 to –6.4. Thus, for the three measures that count some or all distortions as correct (PCC-A, PCC-R, and PPC-R), the NSA/SD group *z* scores are generally 1–3 standard deviation units below the NSA scores, whereas the SD group *z* scores are from 4 to 8

Figure 5. Construct validity data for Normal Speech Acquisition/Speech Delay (NSA/SD) as intermediate to Normal Speech Acquisition (NSA) and Speech Delay (SD) classifications in the SDCS. Box plots indicate the 50th percentile (horizontal bar), 25th and 75th percentiles (box), and 10th and 90th percentiles (vertical lines) for each classification category.



standard deviation units below the NSA scores.

The differences in scores among the NSA, NSA/SD, and SD groups are viewed as strong support for the construct of speech as a continuous trait during this developmental period. NSA/SD is the appropriate classification for children who truly are intermediate between the clinical dichotomy of normal speech versus speech delay. Moreover, in the present context, the differences in scores among the three groups support the use of z score and standard score transformations using the procedures recommended in Appendix B. However, as concluded in the companion article (Shriberg, Austin, et al., 1997), use of z scores and standard scores is recommended only when raw scores are not appropriate for an applied or research question. The reference data in this report provide only preliminary guidance for these needs. Researchers requiring well-standardized normative data on any of the 10 measures should attempt to collect demographically appropriate samples for specific needs.

Acknowledgments

Preparation of this article was supported by research grants R01 DC00496-08 and R01 DC00528-07 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We thank the following people who provided research expertise, research assistance, and/or thoughtful discussion about measurement issues associated with the PCC and the SDCS at different stages of this project: Paula Buckwalter, Michael Chial, Mary Elbert, Peter Flipsen Jr., Lisa Freebairn, Frederic Gruber, Barbara Hodson, Doris Kistler, Joan Kwiatkowski, Gregory Lof, Julie Masterson, Karen Pollock, Bruce Pennington, Carmen Rasmussen, Nancy Records, Dorothy Ross, Hollis Scarborough, Nicholas Schork, Bruce Tomblin, Carol Widder, and Xuyang Zhang.

References

Bernthal, J. E., & Bankson, N. W. (Eds.). (1994). *Child phonology: Characteristics, assessment, and intervention with special populations*. New York: Thieme Medical Publishers.

D'Agostino, R. B., & Stephens, A. (Eds.). (1986). *Goodness-of-fit techniques*. New York: M. Dekker.

Lewis, B. A., & Shriberg, L. D. (1994, November). *Life span interrelationships among speech, prosody-voice, and nontraditional phonological measures*. Miniseminar presented at the Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA.

Shriberg, L. D. (1980). Developmental phonological disorders. In T. J. Hixon, L. D. Shriberg, & J. S. Saxman (Eds.), *Introduction to communicative disorders* (pp. 262–309). Englewood Cliffs, NJ: Prentice Hall.

Shriberg, L. D. (1983). The method of natural phonological processes. In W. Perkins (Ed.), *Current therapy of communication disorders*. New York: Thieme-Stratton.

Shriberg, L. D. (1986). *PEPPER: Programs to examine phonetic and phonologic evaluation records*. Hillsdale, NJ: Lawrence Erlbaum.

Shriberg, L. D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research, 36*, 105–140.

Shriberg, L. D. (1994). Five subtypes of developmental phonological disorders. *Clinics in Communication Disorders, 4*(1), 38–53.

Shriberg, L. D. (1997). Developmental phonological disorder(s): One or many? In B. W. Hodson & M. L. Edwards (Eds.), *Applied phonology: Perspectives and clinical applications*. Gaithersburg, MD: Aspen.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997a). Developmental apraxia of speech: II. Toward a diagnostic marker. *Journal of Speech, Language, and Hearing Research, 40*, 286–312.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997b). Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research, 40*, 313–337.

Shriberg, L. D., & Austin, D. (in press). Comorbidity of speech-language disorder: Implications for a phenotype marker for speech delay. In R. Paul (Ed.), *The speech/language connection*. Baltimore, MD: Paul H. Brookes.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The Percentage of Consonants Correct (PCC) metric: Extensions and reliability data. *Journal of Speech and Hearing Research*, *40*, 708–722.

Shriberg, L. D., Gruber, F. A., & Kwiatkowski, J. (1994). Developmental phonological disorders III: Long-term speech-sound normalization. *Journal of Speech and Hearing Research, 37*, 1151–1177.

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

Shriberg, L. D., & Kent, R. D. (1995). *Clinical phonetics* (2nd ed.). Boston: Allyn & Bacon.

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). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders, 47*, 256–270.

Shriberg, L. D., & Kwiatkowski, J. (1983). Computer-assisted Natural Process Analysis (NPA): Recent issues and data. In J. Locke (Ed.), Assessing and treating phonological disorders: Current approaches. *Seminars in Speech and Language, 4*, 389–406.

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. (1994). Developmental phonological disorders I: A clinical profile. *Journal of Speech and Hearing Research, 37*, 1100–1126.

Shriberg, L. D., Kwiatkowski, J., & Gruber, F. A. (1994). Developmental phonological disorders II: Short-term speech-sound normalization. *Journal of Speech and Hearing Research, 37*, 1127–1150.

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

Received March 8, 1996

Accepted November 29, 1996

Contact author: Lawrence D. Shriberg, PhD, The Phonology Project, Waisman Center on Mental Retardation and Human Development, University of Wisconsin–Madison, 1500 Highland Avenue, Madison, WI 53705. Email: shriberg@waisman.wisc.edu

Copies of the lifespan reference data and additional information are available at the Phonology Project Web site: http://www.waisman.wisc.edu/phonology/

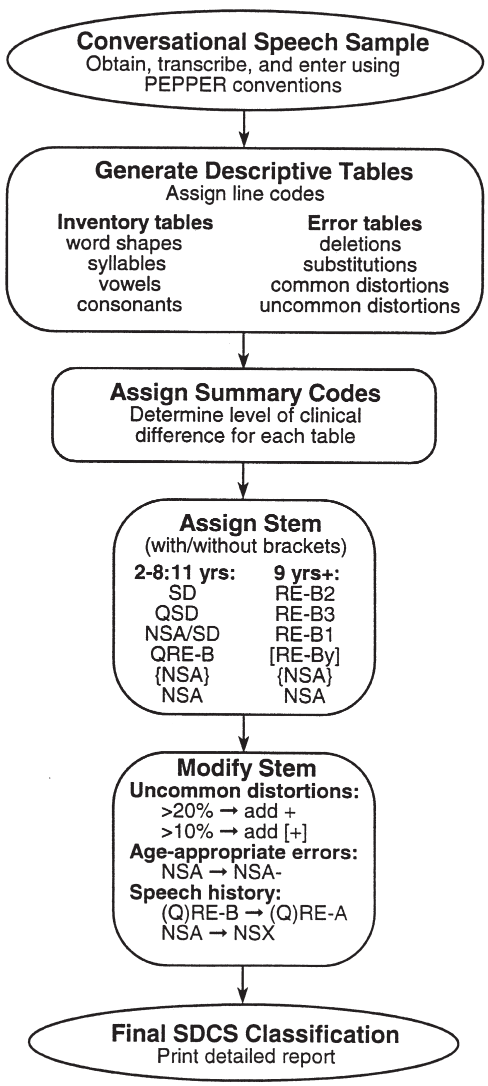
Appendix A: The Speech Disorders Classification System (SDCS).

Colleagues have inquired about technical details of the SDCS program, which will be ported from the VAX environment to a Windows95 PC program. For the interested reader, Figure A1 is a flow chart of the process the SDCS module of PEPPER uses to assign a classification to a conversational speech sample.

Conversational Speech Sample

The program begins with a narrowly transcribed conversational speech sample (see Shriberg, 1993, Appendix, for detailed discussion). The length of the conversational speech sample is important. To address validity and reliability concerns, the SDCS requires at least two tokens in two different words of any error or inventory item to be used in making a classification. As a result, a conversational speech sample rich in different word types yields more items for use in making an SDCS classification. Our lab has used 100 word types (i.e., first-occurrence words that are intelligible and not excluded for other reasons) as a goal for samples to be used for SDCS classification, although this criterion cannot always be met. For this reason, it is important to examine the descriptive tables produced by the SDCS, together with the final classification, to confirm results. The descriptive tables may also be helpful in describing the speech of children who, for reasons discussed in the companion paper (Shriberg, Austin, et al., 1997), cannot provide an adequate conversational speech sample for classification.

Figure A1. Flow chart of the Speech Disorders Classification System (SDCS).



The conversational speech sample is entered into PEPPER using the transcription rules established for other measures, including rules about which words/utterances to disregard, how to handle unintelligible words or parts of words, and other conventions originally described in Shriberg (1986). Whereas other PEPPER analyses (such as the PCC) examine all sounds in a sample, SDCS analysis is selective in the errors that it examines to make a classification, excluding casual word forms, errors that occur only once, word-medial errors, and (most importantly) age-appropriate errors from the classification process. To ensure that casual speech forms are not considered errors, the SDCS consults a table of common words and phrases to be ignored (e.g., variants of "I don't know," "yes," "no," etc.). Errors that occur only once are excluded for reliability concerns, and errors on medial consonants are excluded because there are few reliable developmental data on word-medial consonants.

The SDCS consults several tables in the analysis process to determine whether a sound or word-shape is expected at that age and whether an error is age-appropriate. A detailed discussion of the age-appropriate criteria for normal speech acquisition for ages 2 through 9+ is included in the Appendix of Shriberg, 1993. To maximize sensitivity to some speech problems, a few common and uncommon clinical distortions are considered errors by the SDCS, even when transcribed as nonerrors.

Generate Descriptive Tables

The next step in the SDCS analysis is to generate descriptive tables, including four *inventory tables* (Word Shape Inventory, Syllable Inventory, Vowel Inventory, Consonant Inventory) and four *error tables* (Consonant Deletions, Consonant Substitutions, Common Clinical Distortions, Uncommon Clinical Distortions), which provide item-by-item details about a speaker's inventory of sounds and word shapes and a listing of all errors by type in the sample. For each line in each table, the program provides a code for the speaker's current age (in years), for 1 and 2 years younger, and for 1 year older. To code an individual line in a table, the program consults the appropriate reference tables described above.

Inventory Tables

For the inventory tables, each incorrect and correct occurrence of a sound or word shape is counted for all words and for all different (first-occurrence) words. Only substitution and deletion errors are considered incorrect for the inventories. The Consonant Inventory provides a further breakdown of these counts by initial, medial, and final position, and by singletons and clusters. However, only different word occurrences (in any position, singleton or cluster) are used to determine if the item is *in the inventory* or if it meets *error* criteria. A sound or word shape is determined to be in the inventory if it is produced correctly twice in two different words or produced once correctly without other errors for that sound or word shape. An inventory item meets error criteria if it is produced incorrectly twice (in two different words) and correctly only one or fewer times. An item meets *questionable error* criteria if there is one incorrect production and one or fewer correct productions, or if the sound or word shape is expected to be present at the speaker's age and does not occur in the sample (and is not considered rare).

Error Tables

For the error tables, each deletion, substitution, and distortion (common and uncommon) is listed in the appropriate table. As in the Consonant Inventory, deletions and substitutions are reported by word position and by singleton versus cluster; however, only the initial and final positions are used to determine the line code (i.e., medial errors are disregarded). If more than one substitution type occurs for a consonant (e.g., f/θ and t/θ both occur) the counts for each substitution type appear as separate lines in the Consonant Substitutions table, and the sum of all substitutions for the target sound appear in a single summary line. Only the summary line is used when determining the Substitutions code at the table level.

For 6- to 8-year-old children, substitutions that are perceptually and productively similar to the common clinical distortions (i.e., w/r is similar to derhotacized /r/) are considered only for making a QRE classification and, therefore, are marked differently than other substitution errors. (QRE substitutions are: initial position: w/r, w/l; final position: vowel/r, vowel/l; any position: t/θ, θ/s, ð/z, vowel/ɚ, vowel/ɝ.) A deletion, substitution or distortion meets *error* criteria if it occurs at least twice in two different words and meets *questionable error* criteria if it occurs only once or in only one word. Distortion errors meet *multiple error* criteria if they occur four or more times in four or more different words.

Assign Summary Codes

Next, a summary code is assigned for each of the inventory and error tables. For the inventory tables, a summary code is assigned for the Vowel Inventory, Consonant Inventory, and the combined Word Shape Inventory and Syllable Inventory tables. For the error tables, summary codes are assigned for Deletions, Substitutions, and Common Clinical Distortions, with a separate code assigned for QRE Substitutions. Common clinical distortions are divided into six classes: labialized, dentalized, lateralized, velarized, derhotacized, and epenthetic stop /g/ added to /ŋ/ (i.e., [ŋᵍ]). Each class receives a summary code (as above), and the Common Clinical Distortions summary code is assigned the lowest code for any class. Uncommon Clinical Distortions are considered at a later point in the classification process. As with the line codes, summary codes are assigned for the speaker's current age, 1 and 2 years younger, and 1 year older. Whereas the line codes indicate whether a speaker's productions meet error criteria, the summary codes indicate whether the speaker has a *clinical difference* (CD) in one or more areas of speech production. Criteria for the four levels of clinical difference used to make SDCS classifications are defined in Table A1.

Assign a Stem

After all of the summary codes and the combined codes are assigned, the SDCS consults the classification table that lists the coding criteria for each of the stems and stem/bracket classification categories. Whereas NSA is appropriate for all ages, the remaining classifications are divided into two groups: those appropriate for speakers age 2 to 8 years (NSA, NSA/SD, QSD, SD, QRE) and those appropriate for speakers age 9 years and older (RE-1, RE-2, RE-3). Table A2 lists these classifications with their definitions and coding criteria.

Modify Classification Stems

The last step in the SDCS is to modify the assigned stem/brackets to reflect uncommon distortion errors, age-appropriate errors, and (for NSA and RE classifications) to reflect a speaker's speech history.

Uncommon Clinical Distortions

Because there are few normative data for uncommon clinical distortions, the SDCS does not make classification decisions based on these errors. However, their presence or absence may be important for clinical and/or research questions. To reflect the frequency of occurrence of these errors, a "+" is added to any stem if uncommon clinical distortions occur in more than 20% of words in the sample, and a "[+]" is added if 10–20% of the words have these errors.

Age-Appropriate Errors

For speakers classified as NSA, {NSA}, or NSA/SD, if age-appropriate errors occur a "-" is added to the NSA stem (i.e., NSA-, {NSA-}, NSA-/SD).

Speech History

Finally, the database is examined to determine if the speaker has a history of speech disorders. If the speaker has a previous sample in the database that meets criteria for any of the disorder classifications, or if a database variable indicates a documented history of speech disorder, NSA, QRE, and RE stems are modified to reflect that history. The NSA stem is changed to NSX, where X is the age at which speech normalized (e.g., a child classified as QSD at age 3 and {NSA-} at age 4 would receive a final classification of {NS4-}). The "B" in a QRE-B or RE-B stem is changed to an

Table A1. Criteria for levels of clinical difference (CD) used to make SDCS slassifications.

Level of clinical difference		Criteria	
No.	**Term**	**Inventory Tables**	**Error Tables**
1	Questionable clinical difference	Questionable errors occurred for one or more inventory items.	Questionable errors were found for one or more target consonants.
2	Marginal clinical difference	One sound or word shape meets error criteria.	One target consonant meets error criteria.
3	Clinical difference	Two or more sounds or word shapes meet error criteria.	*Substitutions and Deletions:* Two or three target consonants meet error criteria. *Distortions:* Two or more sounds in at least one class meet error criteria *or* one sound in at least one class meets multiple error criteria.
4	Multiple clinical differences	Not applicable	*Substitutions and Deletions:* Four or more target consonants meet error criteria.

Table A2. Description of and criteria for SDCS classifications.

Eligible ages	Classification	Description	Criteria
All ages	NSA	Normal Speech Acquisition	May include questionable CD.[a] Intelligibility Index (II) > 80%
	{NSA}	Marginal NSA	*All ages*: Marginal CD for Substitutions or Deletions *and/or* CD or marginal CD for any inventory table. *Ages 6–8*: As above *or* marginal CD for Distortions and/or QRE Substitutions
2:0–8:11	NSA/SD	Between NSA and Speech Delay	CD for Vowel Inventory, Deletions, or Substitutions *or* Marginal CD or CD for two of these three tables *or* II between 76% and 80%
2:0–8:11	SD	Speech Delay	Multiple CD for Deletions or Substitutions *or* CD for Vowel Inventory, Deletions, or Substitutions (2 of these 3)
2:0–5:11	QSD	Questionable Speech Delay	Meets criteria for Speech Delay at current age, but does not meet criteria for one year younger
6:0–8:11	QRE	Questionable Residual Errors	CD for QRE Substitutions and/or Distortions
9+	RE-1	Residual Errors 1 - Common Clinical Distortions	CD for Distortions
	RE-2	Residual Errors 2 - Common Clinical Distortions and Imprecise Speech	CD for Distortions *and* Multiple CD for Deletions or Substitutions *or* CD for Vowel Inventory, Deletions, or Substitutions (2 of these 3)
	RE-3	Residual Errors 3 - Imprecise Speech	Multiple CD for Deletions or Substitutions *or* CD for Vowel Inventory, Deletions, or Substitutions (2 of these 3)
	[RE-y], y = 1, 2, or 3	Residual Errors 1, 2, or 3 - with marginal Common Clinical Distortions and/or marginally Imprecise Speech	Marginal CD for Distortions *and/or* marginal CD for Vowel Inventory, Deletions, or Substitutions. Meets other criteria for RE-1, RE-2, or RE-3.

[a] CD: Clinical Difference (see Table A1)

"A" (e.g., QRE-A, RE-A1). If there is no history of speech disorder recorded in the database, the "B" assigned as a default remains in the RE classification (e.g., QRE-B, RE-B1).

The final output of the SDCS program is a multiple page report that includes the inventory tables, error tables, lists of all summary and combined codes, and the SDCS classification. To complete the description of the sample, a panel of all 10 of the measures described in the companion paper (including early, middle, and late subscales where appropriate; Shriberg, Austin, et al., 1997) are printed for the sample being classified and for any earlier or later samples available in the database from the same speaker.

JSLHR, Volume 40, 723–740, August 1997

Appendix B: Statistical analyses of the reference data.

Three questions about the psychometric characteristics of the reference data were posed and answered using descriptive tallies and displays and inferential statistical analyses. Findings were used to develop the tabular formats for the reference data and to guide recommendations in the text for using this information.

Are the Distributions of Scores on the 10 Speech Measures Appropriate for Parametric Statistics on Untransformed (Raw) Scores?

Five descriptive statistics (mean, median, standard deviation, skew, kurtosis) were obtained for each of the 10 speech measures for each Age x Gender subgroup of 3- to 8-year-old NSA children. These statistics were inspected to determine whether untransformed raw scores were appropriate for parametric descriptions and applications, including *z* scores and standard scores. Anderson-Darling tests for normality, which are appropriate for sample sizes above 8 subjects (D'Agostino & Stephens, 1986), indicated that distributions did not statistically differ from normality ($p > .05$) for 34 (34%) of the 100 tests (10 Measures x 5 Ages x 2 Genders; tests could not be performed for the 8-year-old boys and girls in the original analyses because cell sizes were fewer than 8 speakers). As expected, significant departures from normality occurred on the measures for which means approached 100% at the youngest ages, including the Intelligibility Index, PVC, PVC-R, and PCI; only 7 (17.5%) of the 40 statistics for these measures failed to reject normality. The ACI distributions also had significant departures from normality, as might be expected given the nonlinearity of the measure discussed in the companion paper (Shriberg, Austin, et al., 1997).

For the three measures of consonant production, PCC, PCC-A, and PCC-R, however, 16 (53.3%) of the 30 tests failed to reject a null hypothesis of normality, with associated skew and kurtosis values only infrequently above 2.0. For the PPC and PPC-R, which reflect both consonant and vowel production, 9 (45%) of the 20 tests failed to reject normality. Finally, for Anderson-Darling tests on all measures for Age x Gender children in the QRE, NSA/SD, and SD subgroups, 99 (69.7%) of the 140 tests failed to reach the value needed to reject normality.

Departures from patently normal distributions on the 10 quantitative measures were expected because, based on qualitative criteria, the lack of age-inappropriate speech errors in these samples allowed them to be classified by the SDCS as NSA. This effectively skewed scores for some measures towards 100% competence and reduced their range in comparison to distributions on measures that include all performance levels (i.e., both affected and nonaffected individuals). As indicated in the analyses above, however, the primary sense of the data is that departures from normality in the NSA speakers were not so severe as to recommend use of a transformation to improve normality. Moreover, departures from normality within the NSA subjects occurred least for Age x Gender groups on the PCC, PCC-A, and PCC-R. Together with the evidence from the descriptive statistics illustrated in the text—particularly the sensitivity of the PCC, PCC-A, and PCC-R measures to differentiate SDCS classification subgroups—the outcomes from these analyses were viewed as providing statistical support for the use of parametric procedures with the reference data for 3- to 8-year-old children. Specifically, these analyses supported the use of untransformed raw scores and the use of means and standard deviations of the raw scores as measures of central tendency and dispersion, respectively.

At 9 years and above, NSA speakers' scores approach 100% on all measures. Therefore, the data are too skewed to justify use of a transformation in order to use standard deviations to compute Age x Gender *z* scores. However, as indicated in the text, means and standard deviations are used for descriptive purposes because medians on each measure would invariably be an uninformative 100%. To aid in the descriptive value of these data for older speakers, the reference data include the lowest score obtained on the measure by speakers in each SDCS Age x Gender subgroup.

Do the Descriptive Statistics Support the Use of Age x Gender Statistics to Compute Standard Scores?

Following the positive conclusion on the suitability of the 3- to 8-year-old data for parametric procedures, the next analyses examined mean and standard deviation trends for the NSA speakers in these Age x Gender subgroups. The goal was to determine if it was appropriate to use means and standard deviations for the purposes of *z* score and standard score calculations.

Age Trends for Standard Deviations

The PCC-R was selected as the most appropriate measure on which to test trends within and across different age and SDCS groups. Standard deviations of the NSA Age x Gender subgroups on the PCC-R varied from 3.1 to 5.3 in the 3- to 5-year-old groups and from 1.0 to 3.1 for the 6- to 8-year-old groups. Although these differences in standard deviations are minor in absolute magnitude, the ratio of differences among standard deviations (as large as 3:1) would have important consequences for standard scores (which are most appropriate when subgroups have approximately similar standard deviations).

To test an alternative approach, means and standard deviations in the 3- to 5-year-old groups and 6- to 8-year-old groups were compared to determine if it might be useful to constitute two larger groups with common standard deviations. Because girls scored slightly higher than boys for most age groups, collapse across gender was not considered. These two age groupings were considered appropriate because they share the same potential SDCS classifications (3–5: NSA, NSA/SD, QSD, SD; 6–8: NSA, NSA/SD, SD, QRE). Moreover, although allowable errors change somewhat from year to year, the largest change in expectations occurs after the age of 5.

For the 3- to 5-year-old children of each gender, *t* tests were performed that compared one age group to the two remaining age groups (3 vs. 4 and 5, 4 vs. 3 and 5, and 5 vs. 3 and 4). Results differed by speech measure, with two subgroups of measures differentiated by whether or not they counted all distortions (PPC-R, PCC-R, PVC-R) or only uncommon clinical distortions as errors (PCC-A), versus the measures that count all clinical distortions as errors (PPC, PCC, PVC). For the former group of four measures, the only statistically significant difference between age groups ($p < .05$) was the *t* test for 5-year-old girls compared to the 3- and 4-year-old girls on PCC-A scores. For each of the measures in the latter subgroup (PPC, PCC, PVC), most comparisons of performance at age 3 versus 4/5, 4 versus 3/5, and 5 versus 3/4 were statistically significant at alpha levels ranging from .05 to .001. These findings indicated that the 3- to 5-year-old children could be considered as one group for those measures that scored all or some distortions as correct. However, because scores on the measures that score distortions as incorrect differ statistically by age, caution should be used when using the grouped reference data for these measures.

In like manner, *t* tests were performed for 6- to 8-year-old groups (6 vs. 7 and 8, etc.). Statistically significant differences were found for most comparisons, except for the two vowel measures (PVC and PVC-R). On inspection, the large number of significant differences appeared to be associated with disproportionate cell sizes. The conclusion from these analyses was that it was also appropriate to derive a common standard deviation for each gender in these older children.

Procedure to Derive Common Standard Deviations

Combined standard deviations, one for each gender for all 3- to 5-year-olds and one for each gender for all 6- to 8-year-olds, could be calculated in two ways. One way would be to ignore age within subgroups and calculate standard deviations for each gender in each subgroup (3–5 and 6–8), which would preserve whatever weighting was associated with each age (i.e., recall that the cell sizes differ considerably per Age x Gender subgroup). Alternatively, the standard deviations at each age within subgroups could be averaged, which would give equal weight to each age group.

As each of the above procedures could be justified from the perspective of at least one of the Age x Gender groups, the decision was to combine the two approaches. For the 3- to 5-year-olds, the standard deviations for boys and girls were calculated from the scores pooled over age, and the standard deviations for each age group were averaged. These two scores were then averaged to determine a *derived* standard deviation recommended for use to calculate *z* scores and standard scores (see below).

Similar procedures were used to derive a standard deviation for NSA children in the 6- to 8-year-old group. Because the 7- and 8-year-old groups were significantly smaller in size than the 6-year-old group, with standard deviations as small as 1.0, an average of these standard deviations might have produced *z* scores that presented a false picture of severity of involvement for the 6-year-olds. As an alternative, the pooled standard deviations were averaged with the 6-year-old standard deviations to calculate the derived standard deviation.

Thus, as shown in the sample reference data in Table 2, two standard deviations are calculated for the NSA boys and girls. The first SD (*SD1*) is the standard deviation of the Age x Gender x SDCS subgroup. The second SD (*SD2*) is the derived standard deviation. The reference data also include means and standard deviations from the two larger age groups (3–5 and 6–8) used to calculate the derived standard deviations.

What Is the Most Appropriate Procedure to Derive *z* Scores and a Standard Score for the Reference Data?

The third question concerned procedures to transform raw scores to *z* scores, and, in turn, *z* scores to standard scores.

Z Scores

As described in the text, when raw scores for the 10 measures are used descriptively, it is appropriate to use the standard deviations obtained for each measure on each Age x Gender subgroup. However, when *z* scores are used to compare children of different ages and gender, a different approach is recommended based on the preceding analyses.

A *z* score for a measure is computed by dividing the difference between the reference mean and an individual's score by the reference standard deviation. On the basis of the above analyses, a combined approach is suggested for obtaining *z* scores from the reference data. For the reference mean term in the calculation, the mean of a subject's appropriate Age x Gender subgroup should be used. However, for a maximally conservative comparison of *z* scores for children of widely different ages, the standard deviation should be selected from the appropriate grouped data (i.e., the *derived* standard deviations (*SD2*) for each gender for 3- to 5-year-olds and 6- to 8-year-olds). Again, use of the derived standard deviation negates the influence of differences in standard deviations associated with cell-size differences in the reference data.

Standard Scores

As described in the text, *z* scores are often transformed into an alternative standard score solely for ease of understanding. *Z* scores typically contain decimal points, may be negative numbers, and their range is not intuitively understood. Standard scores meet these needs by setting a mean to a more intuitive number (e.g., 50 or 100) and by selecting a standard deviation that best accommodates the range of expected *z* scores. Standard scores are calculated by multiplying a subject's *z* score by the chosen standard deviation and adding or subtracting (based on the sign of the *z* score) the product from the chosen mean.

The choice of a mean and standard deviation for the standard score for the present data was influenced by two factors. First, a mean of 100 was perceived as more intuitive than a mean of 50. Second, because the raw scores of speech-delayed children were so much lower than those of the

NSA children on some measures, the standard deviation for the standard score would need to be relatively small. For example, the SD children scored significantly below the NSA children on the PCC-R, with scores ranging from 32.0% to 90.7% for 3- to 5-year-old children, and 47.4% to 91.5% for 6- to 8-year-old children. *Z* score transformation of the lowest score in the 6- to 8-year-old group (47.4%) using the derived standard deviation of 3.0 resulted in a *z* score of –15.6 for a 6-year-old male (i.e., 15.6 standard deviations below the mean of 6-year-old NSA boys).

The above considerations provided rationale for setting the mean for the standard score to 100 and the standard deviation to 5. Using these values, the *z* score of –15.6 above transforms to a standard score of 22 (i.e., remains a positive standard score). Thus, NSA children with positive *z* scores will have standard scores above 100 (three standard deviations above = 115), whereas samples from very involved children might approach zero on some measures. For measures with very high subgroup means and small standard deviations (PCI, PVC-R, II and Early-8 subscales on the PCC, PCC-A, PCC-R, and PCI), some very severely involved SD children may have *z* scores that transform to a negative standard score.