

Supplement Article

A Diagnostic Marker to Discriminate Childhood Apraxia of Speech From Speech Delay: II. Validity Studies of the Pause Marker

Lawrence D. Shriberg,^a Edythe A. Strand,^b Marios Fourakis,^a Kathy J. Jakielski,^c Sheryl D. Hall,^a Heather B. Karlsson,^a Heather L. Mable,^a Jane L. McSweeney,^a Christie M. Tilkens,^a and David L. Wilson^a

Purpose: The purpose of this 2nd article in this supplement is to report validity support findings for the Pause Marker (PM), a proposed single-sign diagnostic marker of childhood apraxia of speech (CAS).

Method: PM scores and additional perceptual and acoustic measures were obtained from 296 participants in cohorts with idiopathic and neurogenetic CAS, adult-onset apraxia of speech and primary progressive apraxia of speech, and idiopathic speech delay.

Results: Adjusted for questionable specificity disagreements with a pediatric Mayo Clinic diagnostic standard, the estimated sensitivity and specificity, respectively, of the PM were 86.8% and 100% for the CAS cohort, yielding positive and negative likelihood ratios of 56.45 (95% confidence interval [CI]: [1.15, 2763.31]) and 0.13 (95% CI [0.06, 0.30]). Specificity of the PM for 4 cohorts totaling 205 participants with speech delay was 98.5%.

Conclusion: These findings are interpreted as providing support for the PM as a near-conclusive diagnostic marker of CAS.

The first article in this series (PM I; Shriberg et al., 2017a) includes a review of the development of a diagnostic marker of childhood apraxia of speech (CAS) termed the Pause Marker (PM). The goal was to develop a procedure that maximally meets the seven attributes of a highly valued diagnostic marker described and prioritized in that article's Table 1. In addition to the requirements of high diagnostic accuracy and high test-retest reliability, the primary goal in developing the marker was to base it on the occurrence of a quantified behavioral event that has strong theoretical coherence with the neurocognitive and neuromotor substrates of CAS.

To briefly review the PM procedure, a behavioral correlate of CAS termed an *inappropriate between-words pause* is proposed to be consistent with the "moment" of apraxia. Identification and computer processing of eight

types of inappropriate between-words pauses in continuous speech are completed using auditory and visual perceptual procedures and software routines for narrow phonetic transcription, prosody-voice coding, and acoustics analyses—all in a platform called Programs to Examine Phonologic and Phonetic Evaluation Records (Shriberg, Allen, McSweeney, & Wilson, 2001). The software computes a speaker's PM score by dividing the total number of four types of inappropriate between-words pauses in continuous speech (termed *Type I pauses*) by the number of opportunities for between-words pauses in the sample and multiplying the quotient by 100. As with other measures in the assessment protocol to be described, the resulting percentage of inappropriate between-words pauses is subtracted from 100 so that higher PM scores indicate higher performance. PM scores calculated on fewer than 40 opportunities for between-words pauses are classified as *indeterminate*, with the recommendation that another speech sample be collected on the same or a later date to include the required number of pause opportunities or be merged with the prior sample to meet the pause-opportunity criterion.

A cutoff percentage score identified in PM developmental studies is used to classify participants as *PM+* (positive for CAS) or *PM-* (negative for CAS). PM scores

^aWaisman Center, University of Wisconsin–Madison

^bDepartment of Neurology, Mayo Clinic, Rochester, MN

^cDepartment of Communication Sciences and Disorders, Augustana College, Rock Island, IL

Correspondence to Lawrence D. Shriberg: shriberg@waisman.wisc.edu

Editor and Associate Editor: Julie Liss

Received August 24, 2016

Revision received April 12, 2016

Accepted August 21, 2016

https://doi.org/10.1044/2016_JSLHR-S-15-0297

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

within 1 percentage point lower or higher than the cutoff percentage, termed *marginal PM scores*, are resolved if possible by using findings for three supplemental signs of CAS: slow articulatory rate (Rate), inappropriate sentential stress (Stress), and transcoding errors (Transcoding). Standardized (*Z* deviates) Rate scores are obtained using acoustics-aided auditory perceptual information from the continuous speech sample used for the PM and a reference database of typical same-sex speakers using nonverbal ages when available. The criterion-referenced cutoff for Stress is a Prosody-Voice Screening Profile (Shriberg, Kwiatkowski, & Rasmussen, 1990) Stress score below 80%. Criterion-referenced Transcoding information is obtained from the Syllable Repetition Task (Shriberg et al., 2009; Shriberg & Lohmeier, 2008). A Phonology Project technical report (Tilkens et al., 2017) includes additional PM findings and audio exemplars of inappropriate between-words pauses, including examples of the primary type of inappropriate between-words pauses, termed an *abrupt between-words pause* (see PM III; Shriberg et al., 2017b).

Statement of the Problem

PM I proposed a rationale for a diagnostic marker of CAS that meets a number of measurement goals. The three questions that follow comprise the goals of, respectively, the present, third, and fourth articles in this supplement:

1. PM II: Question 1. Do findings from construct and criterion validity studies support the diagnostic accuracy of the PM to discriminate CAS from speech delay (SD)?
2. PM III: Question 2. Do findings from the PM and other measures support the hypothesis of core representational and transcoding deficits in CAS, and is the PM theoretically coherent with those deficits?
3. PM IV (Shriberg et al., 2017c): Question 3. Do findings from cross-sectional and retrospective longitudinal case studies support an ordinal scale of PM scores, termed the Pause Marker Index, to quantify the severity of CAS for research and clinical applications?

Method

Participants

Eligibility for PM Scoring

Table 1 is a summary of sociodemographic and speech information for 296 of an original group of 320 participants assessed for this project. The 320 original participants were assessed in collaborative studies at several university and hospital settings over the past several decades. The demographic and speech information in Table 1 is from the 296 participants (92.5% of the original sample) whose speech samples met continuous-speech sampling criteria for the

PM scores used to classify each participant as positive or negative for apraxia of speech and who had Syllable Repetition Task scores available in their case records. As described previously, continuous speech samples used to derive PM scores must include at least 40 between-words opportunities for inappropriate pauses. Participants with very low mean utterance lengths (one- to two-word utterances) and/or extremely low intelligibility may not produce continuous speech samples that meet this requirement. Thus, as with other studies requiring behavioral responses to tasks assessing the behavior of interest, findings from this study cannot be generalized to speakers with very low verbal output. The highest number of indeterminate PM scores for participants with SD (i.e., marginal PM scores that required resolution by scores on the three measures described in PM I) were those due to the use of retrospective data that did not include scores on the Syllable Repetition Task, which was developed years after many of the participants with SD had been assessed.

The 296 participants whose data appear in Table 1 were from three groups: children and adults suspected to have active or persistent CAS ($n = 60$), adults meeting Mayo Clinic System (MCS; Shriberg, Potter, & Strand, 2011) criteria for one of two types of adult-onset apraxia of speech (AAS) described later ($n = 31$), and children meeting the Speech Disorders Classification System (SDCS; Shriberg, Strand, & Mabie, 2017) criteria for SD (see PM I; $n = 205$).

We use the generic term apraxia of speech for all forms of this motor speech disorder, using other terms (i.e., CAS, AAS, PPAOS) for different specific clinical entities. Notice that the term CAS is used for both children and for persistent CAS in adults.

Participants in each of the three groups were assessed using current and prior versions of a speech-language assessment protocol described later in this supplement article. The inclusion criteria for each of the eight study cohorts required English as the only or primary language spoken in the home and the intellectual ability and psychosocial disposition to complete the 2-hr speech-language assessment protocol. As described later in this supplement article, the available participant data from the protocol ranged from one continuous-speech sample for some of the earliest assessed participants to responses for all speech-language tasks in the current protocol. The PM uses information from only the continuous-speech sample. Data from other tasks were used in supplemental PM information and in this and other validation studies in this series to be described later. All participants and/or their parents signed assent and consent forms approved by local institutional review boards.

Idiopathic and Neurogenetic CAS

The 60 participants who were suspected to be positive for CAS were assessed by several examiners. A total of 34 (56.7%) of the participants were assessed for suspected CAS by the second author at the Mayo Clinic Department of Neurology, Rochester, Minnesota, using the protocol

Table 1. Demographic and speech information for a total of 296 of 320 study participants with usable speech samples and determinate Pause Marker (PM) scores.

Group	Cohort	Title	Number of participants		Age (years)			% Male/ % female	PCC (%)		
			Samples usable for PM scores	Indeterminate PM scores	M	SD	Range		M	SD	Range
Suspected CAS	CASI	CASI	41	1	9.0	4.2	4–23	68.3/31.7	76.0	13.3	36.8–98.4
	CASN ^a	CASN	19	1	10.9	5.2	4–25	47.4/52.6	75.8	12.3	44.9–92.2
AAS	AOS	Total	60	2	9.7	4.6	4–25	61.7/38.3	76.0	12.9	36.8–98.4
		AOS	15	1	62.3	11.6	45–82	73.3/26.7	91.6	7.3	68.9–99.4
	PPAOS	PPAOS	16	1	71.0	9.0	53–84	56.3/43.7	93.0	6.2	74.0–97.9
		Total	31	2	66.8	11.1	45–84	64.5/35.5	92.3	6.6	68.9–99.4
SD	Clinical	SD1	82	6	4.4	1.3	3–9	74.4/25.6	73.4	12.6	17.5–99.1
		SD2	22	1	5.5	0.6	5–7	77.3/22.7	82.0	6.9	66.4–91.3
	Research	SD3	72	12	4.0	0.7	3–5	73.6/26.4	70.0	9.6	36.2–87.2
		SD4	29	1	4.5	0.9	3–7	48.3/51.7	68.8	11.4	42.1–82.8
	Total	205	20	4.4	1.1	3–9	70.7/29.3	72.5	11.5	17.5–99.1	
Total all samples			296	24							

Note. Speakers comprise two cohorts of participants who were suspected to have childhood apraxia of speech (CAS)—idiopathic CAS (CASI) and neurogenetic CAS (CASN); two cohorts with adult-onset apraxia of speech (AAS): apraxia of speech (AOS) and primary progressive apraxia of speech (PPAOS); and four cohort samples of participants with speech delay (SD).

PCC = percentage of consonants correct.

^aIncludes participants with copy-number variants ($n = 9$) identified in related research and participants with neurodevelopmental disorders associated with disruptions in *FOXP2* ($n = 2$), 4q;16q translocations ($n = 3$), 16p11.2 microdeletion syndrome ($n = 2$), terminal deletion of chromosome 22 ($n = 1$), Joubert syndrome ($n = 1$), and Prader–Willi syndrome ($n = 1$).

described later in this supplement article. Four (6.7%) of the participants were assessed for suspected CAS by the fourth author at Augustana College, Rock Island, Illinois, using the same protocol. The remaining 22 (36.7%) participants who were suspected to have CAS were assessed by a number of examiners in collaborative genetic, neuroimaging, or speech-sound-disorder classification studies using subsets of tasks from the assessment protocol. Data reduction procedures over several decades were completed primarily by the same research personnel, using computer-aided perceptual and acoustic methods for phonetic transcription, prosody-voice coding, and acoustic analyses (Shriberg et al., 2010a, 2010b).

Diagnostic (i.e., “gold”) standard classification of participants as positive (CAS+) or negative (CAS–) for CAS was completed by the second author, who, as indicated, had administered the assessment protocol to over half of the 60 participants who referral clinicians suspected were positive for CAS. That author used the MCS criteria described in Table 2 of PM I to classify all participants’ CAS status, using the entire digital recording of each participant assessed by the referring clinician and other clinicians. We underscore that, as with all bootstrap approaches to the validation of tests and measures (the PM is a measure or task), estimates of the criterion validity of the PM are constrained by the validity of the diagnostic standard against which its scores are compared.

To estimate the interjudge reliability of the second author’s classifications in a study of CAS in galactosemia, an experienced colleague was trained to use the MCS as depicted in Table 2 of PM I. Interjudge reliability for a sample of 10 participants in the study, including participants the

second author of this supplement article had classified as either positive or negative for CAS, was 90% (Shriberg, et al., 2011). To date, this is the only available estimate of interjudge agreement for the MCS diagnostic criteria for CAS, and it is therefore a psychometric limitation of the diagnostic standard used in this study series. As shown in Table 1, each of the participants classified using the MCS criteria as CAS+ or CAS– was placed in one of two cohorts on the basis of case-history information obtained from the assessment protocol. Participants with no known neurogenetic background were termed *CAS-Idiopathic* (CASI; $n = 41$), and those with documentation of a genetic finding were termed *CAS-Neurogenetic* (CASN; $n = 19$). The latter classifications included participants with copy-number variants identified in associated genetic research ($n = 9$) and participants with neurodevelopmental disorders associated with disruptions in *FOXP2* ($n = 2$), 4q;16q translocations ($n = 3$), 16p11.2 microdeletion syndrome ($n = 2$), terminal deletion of chromosome 22 ($n = 1$), Joubert syndrome ($n = 1$), and Prader–Willi syndrome ($n = 1$). As indicated in Table 1, PM scores were indeterminate from one participant in each of the CAS groups. The CAS groups were 61.7% male, 38.3% female; the CASN percentage of 47.4% male (52.6% female) is lower (higher) than the typically reported percentages of speech sound disorder of approximately 70% male (approximately 30% female). The average percentage of consonants correct (PCC) for the CAS participants was 76% ($SD = 12.9\%$).

AAS

The second two participant cohorts in Table 1, collectively termed *AAS*, were included in the present study to

assess support for the convergent construct validity of the PM to identify apraxia of speech. As proposed in PM I, just as in childhood dysarthria, speech processing correlates of CAS should be at least generally consistent with those documented for the adult-onset disorder of the same name. Differences in time of onset clearly have consequences for pathobiology and associated deficits. The present perspective, however, is that core processes in apraxia of speech (the conventional term for AAS) and CAS (a relatively recent term for a disorder resembling AOS that children are suspected to have) are assumed to be substantially similar.

A total of 31 adult participants who met MCS criteria for AAS were assessed: 15 with AOS primarily consequent to stroke and 16 participants with primary progressive apraxia of speech (PPAOS) associated with disease. Informative descriptions of PPAOS include ones by Duffy et al. (2015), Duffy and Josephs (2012), and Josephs et al. (2006, 2012). All AAS classifications were completed and/or confirmed by the second author or a clinical research colleague at the Mayo Clinic. As indicated in Table 1, PM scores were indeterminate from one participant in each of the AAS groups. As shown in Table 1, the AAS group was 64.5% male, 35.5% female; their average PCC score was 92.3% ($SD = 6.6\%$). Other than the reference data used to standardize measures (described later in this supplement article), from 50 adults with typical speech, the research design did not require a control group of adult speakers without AAS (i.e., AAS-) to assess discriminant construct validity.

During the acoustics-aided PM scoring (described later in this supplement article), nine of the 31 (29%) speakers with AAS (four AOS; five PPAOS) were observed to have a phonatory style in which they maintained near-continuous voicing. This style was incompatible with between-words pausing. The PM technical report includes audio examples of this continuous voicing style. It was not observed in any of the participants who were suspected to be positive for CAS, nor was near-continuous voicing observed in any of the children with SD. Without the availability of speech samples from each speaker prior to the adult onset of apraxia of speech, the history of this behavior in each of the nine participants is unknown. Given its high prevalence in this sample, we speculate that the continuous or near-continuous voicing could be a compensatory behavior to inhibit inappropriate between-words pauses.

SD

Case records and digital audio recordings were assembled for a total of 205 participants who met the SDCS criterion for SD of age-inappropriate speech sound deletions and/or substitutions (Shriberg, 1993). Participants in the four SD cohorts shown in Table 1 were a random sample of participants who had been assessed in the Phonology Clinic at the Waisman Center, University of Wisconsin–Madison (SD1) and three cohorts (SD2–SD4) who had participated in collaborative research conducted in eastern,

midwestern, and western U.S. cities. As described in PM I, and consistent with the title of this supplement, the rationale for including a large number of children with SD was to test the specificity of the PM, because severe SD is the clinical entity most often misdiagnosed as CAS. The goal was to assemble a large, geographically diverse sample that would assess specificity and provide useful per-participant information on potential sources of false positives and false negatives for CAS using the PM.

All participants met conventional inclusion criteria for SD of unknown (idiopathic) origin, including no frank cognitive, structural, sensory, motor, or affective deficit as determined by referral information, case-history questionnaire data, and direct assessment findings. Each of the four SD cohorts included 22 to 82 participants. Methods for data collection were generally similar for all studies, with an examiner administering all or a subset of the speech assessment protocol (see Assessment section).

Participants in the four study cohorts ranged in age from 3 to 9 years ($M = 4.4$; $SD = 1.1$), were approximately 70.7% male and 29.3% female, and had PCC scores ($M = 72.5\%$; $SD = 11.5\%$) consistent with the average PCC scores found in preschool children with SD (Shriberg, 2010). Findings for the sample of children with SD (205/225 = 91.1% usable samples) are also likely representative, considering the conventional inclusion and exclusion criteria used to recruit participants in each of the four SD study cohorts and the size and geographic diversity of the cohorts.

Assessment

All sessions were completed in quiet rooms by examiners experienced in assessing children and adults with communicative disorders. Table 2 contains the titles of the 17 speech assessment tasks administered to some of the participants in each of the three groups described in Table 1, as shown in the footnotes and depending on their age group and study cohort. The left-to-right arrangement of cohorts generally parallels the chronological sequence in which most participants in each group were assessed. For example, audio-recorded conversational speech samples were the only type of speech task available from SD1 participants assessed several decades ago, whereas most participants in the CAS and AAS cohorts were assessed with most of the 17 speech tasks on the Madison Speech Assessment Protocol (MSAP; Shriberg et al., 2010a, 2010b), which provides information on the competence, precision, and stability of a participant's speech, prosody, and voice. The digitally recorded stimuli for the MSAP were presented to participants on laptop computers with external tabletop speakers adjusted for comfortable listening. All recordings of participants' responses used high-quality analog or digital audio recorders with matching external microphones. Analog recordings were digitized, using conventional procedures, for the perceptual and acoustic analyses. The complete MSAP also includes a case-history form and additional tests and tasks that provide information about sociodemographic, cognitive-linguistic, hearing, language, medical,

Table 2. The 17 speech tasks in the Madison Speech Assessment Protocol (MSAP) and the numbers of participants in each cohort for whom information was available for each task.

Task	Protocol				SD cohorts				CAS cohorts		AAS cohorts	
	A ^a	B ^b	C ^c	D ^d	SD1	SD2	SD3	SD4	CASI	CASN	AOS	PPAOS
Conversational Speech Sample	X	X	X	X	88	23	84	30	40	20	16	17
Challenging Words Task	X	X	X	X	0	0	83	30	40	19	16	17
Emphatic Stress Task	X	X	X	X	0	0	69	30	22	19	6	17
Goldman-Fristoe Test of Articulation—Second Edition	X	X	X		0	0	82	30	40	18	0	0
Lexical Stress Task	X	X	X	X	0	0	83	30	22	20	6	17
Multisyllabic Words Task 1	X	X			0	0	0	28	17	16	0	0
Multisyllabic Words Task 2			X	X	0	0	0	0	6	5	6	17
Nonword Repetition Task	X	X	X	X	0	0	84	30	40	19	6	17
Syllable Repetition Task	X	X	X	X	0	23	84	30	40	19	6	17
Sustained Consonant Task	X	X	X	X	0	0	0	27	21	15	5	17
Speech Phrases Task	X	X	X	X	0	0	0	3	22	18	4	15
Sustained Vowel Task	X	X	X	X	0	0	0	29	22	15	6	17
Vowel Task 1	X	X	X	X	0	0	0	30	39	18	6	17
Vowel Task 2	X	X	X	X	0	0	0	30	39	19	6	17
Vowel Task 3		X	X	X	0	0	0	0	15	17	6	17
Rhotics and Sibilants Task		X	X	X	0	0	0	0	16	17	4	17
Diadochokinesis Task	X	X	X	X	0	0	0	24	22	17	6	17

Note. SD = speech delay; CAS = childhood apraxia of speech; AAS = adult-onset apraxia of speech; CASI = idiopathic CAS; CASN = neurogenetic CAS; AOS = apraxia of speech; PPAOS = primary progressive apraxia of speech.

^aPreschool (ages 3.0–5.9 years). ^bSchool age (ages 6.0–11.9 years). ^cAdolescent (ages 12.0–17.9 years). ^dAdult (ages 18.0 years).

educational, speech therapy, and psychosocial status and history.

As described in PM I, the conversational speech samples provided the speech data for the PM and the Rate and Stress signs. The Syllable Repetition Task provided the information for the Transcoding sign (phoneme additions) described in PM I. Reliability estimates for elements of phonetic transcription, prosody-voice coding, and acoustic analyses are also included in PM I.

Results

Three analysis series were completed to answer Question 1 in the statement of the problem: Do findings from construct and criterion validity studies support the diagnostic accuracy of the PM to discriminate CAS from SD? To minimize potential Type II statistical errors associated with the relatively small cell sizes for most of the inferential statistics computed for the analyses to follow, familywise 95% confidence levels were used to test the statistical significance of the effect sizes for between-groups mean differences.

The first analyses to follow, under the CAS Analyses heading, report the percentage of the 60 participants who were suspected to be positive for CAS and were classified as either CAS+ or CAS– by both the MCS and PM criteria. These findings estimate the sensitivity and specificity of the PM, which in turn provides the positive and negative likelihood ratios used to estimate the concurrent validity (i.e., diagnostic accuracy) of the PM. The CAS analyses also examined methodological and participant variables possibly associated with differences in classification outcomes (i.e., disagreements in the classification of

participants as CAS+ or CAS– using the MCS and PM criteria).

The second and third analyses to follow, AAS analyses and SD analyses, report PM diagnostic accuracy findings for, respectively, the 31 AAS participants classified using adult versions of the MCS and the 205 participants classified as having SD using the SDCS. Both analyses also include information on possible sources of disagreement in using the PM versus the MCS criteria for AAS and the SDCS criteria for SD.

CAS Analyses

Figure 1 includes estimates of the diagnostic accuracy of the PM on the basis of MCS and PM classification outcomes for the 60 participants who were suspected to be positive for CAS in Table 1. Figures 1A–1C provide diagnostic accuracy estimates—that is, the percentage of comparisons in which the CAS classification outcome from the PM agrees with the CAS classification outcome from the MCS for participants in, respectively, the CASI group, the CASN group, and the combined CASI and CASN groups. As described next, the specificity findings in the middle bar in each panel include classification outcomes termed “questionable,” whereas the specificity findings in the rightmost bar estimate specificity with the questionable classification outcomes excluded.

Beginning with the CASI sensitivity estimates in Figure 1A, the MCS criteria listed in Table 2 of PM I classified 24 of the 41 participants with suspected CASI as CAS+. The PM score classified 21 of these 24 participants as CAS+, yielding a PM sensitivity estimate of 87.5%. As shown for the specificity estimate in the middle bar of

Figure 1. Diagnostic accuracy estimates for the Pause Marker based on the Mayo Clinic System and Pause Marker classification outcomes for the 60 participants who were suspected to be positive for childhood apraxia of speech (CAS), for participants in (A) the group with indeterminate CAS ($n = 41$), (B) the group with neurogenetic CAS ($n = 19$), and (C) the two groups combined ($n = 60$).

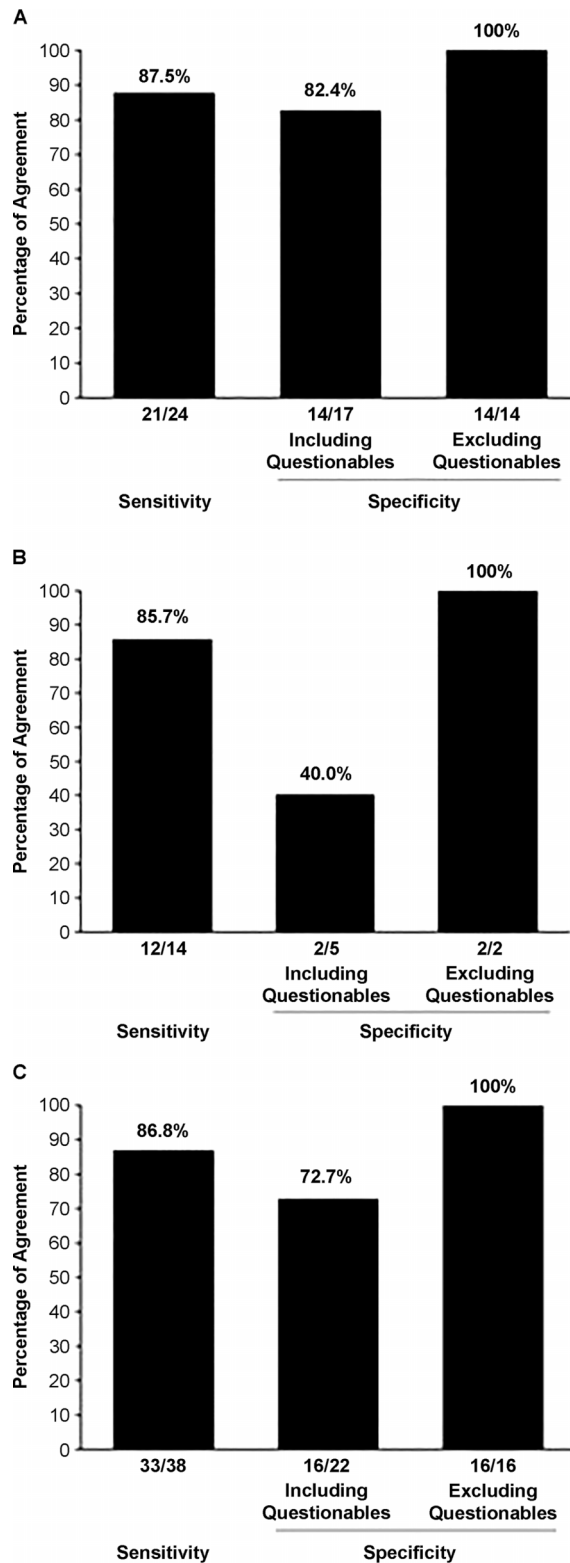


Figure 1A, the MCS classified the 17 remaining participants with suspected CASI as CAS–, whereas the PM classified 14 of these 17 participants as CAS–, yielding an initial PM specificity estimate of 82.4% (middle bar).

The rationale and findings for considering the three MCS/PM differences in classification outcomes as “questionable disagreements” are presented in the following analyses of disagreements. As shown in the rightmost bar in Figure 1A, excluding these three questionable differences from the calculation (i.e., leaving 14 classification agreements) yields a PM specificity value of 100%. Excluding three of the five questionable MCS/PM classifications from the 19 CASN specificity comparisons (see Figure 1B) and six of the 22 MCS/PM classifications provides similar estimates for the CASN participants, and Figure 1C provides similar estimates for the combined CASI and CASN cohorts.

As shown in Figure 1C, the estimated sensitivity and adjusted specificity of the PM for the combined CASI and CASN cohorts are 86.8% and 100.0%, respectively. Using an approach suggested by Haldane to accommodate zero cell constraints (Walter & Cook, 1991), these values yield a statistically significant positive likelihood ratio of 56.45 (95% confidence interval [CI] [1.15, 2763.31]) and a significant negative likelihood ratio of 0.13 (95% CI [0.06, 0.30]). These likelihood-ratio values are close to the suggested requirements for a conclusive diagnostic marker of > 10.0 and < 0.10 , respectively, for positive and negative likelihood ratios (see the discussion of issues associated with confidence intervals in Dollaghan, 2007). As indicated previously, the goals of the detailed examinations of MCS/PM classification outcome disagreements to follow are to motivate the adjusted estimate of the diagnostic accuracy of the PM and to examine participant and speech-sampling variables that might inform diagnostic assessment in CAS using the PM.

Consensus CAS+ and CAS– Groups

To provide a means to examine participant and sampling variables possibly associated with MCS/PM classification disagreements, two consensus CAS groups were assembled. As shown in the first section of Table 3, the Consensus CAS+ group contains information for 31 of the 60 participants classified CAS+ by both the MCS and PM criteria. The Consensus CAS– section of Table 3 contains this information for the 16 participants classified CAS– by both the MCS and PM criteria. The rationale for the two consensus groups is that because they include only participants classified similarly by the two diagnostic approaches, participants in these groups are those most likely to be true positives (Consensus CAS+) and true negatives (Consensus CAS–) for CAS.

The information from the two consensus groups provides a form of reference data against which to examine findings for the 11 participants (see Table 3, Classification disagreements) whose CAS status was classified differently by the PM than by the MCS (as shown in Table 1, two participants had indeterminate PM scores). As shown in Table 3, data for these 11 participants are split into two

groups. Participants 1–5 were classified CAS+ by the MCS criterion (MCS+) and CAS– by the PM score (PM–). Participants 6–11, by contrast, were classified CAS– by the MCS criterion (MCS–) and CAS+ by the PM score (PM+).

A number of variables in Table 3 were examined to determine if averaged findings for the two disagreement groups are significantly different from the Consensus CAS+ or the Consensus CAS– group. The primary question in the following analyses is whether methodological constraints might be sources of the disagreements in classification outcomes using the two CAS diagnostic classification systems (see Figure 1). Disagreements for which a methodological variable may have been the source of the classification disagreement (e.g., insufficient participant assessment data) are termed “questionable disagreements.” As described in the following sections, the five MCS+/PM– disagreements (see Table 3, Participants 1–5) affecting the estimate of the sensitivity of the PM are concluded to be valid disagreements, reflecting the different perspectives on the diagnostic signs proposed to define CAS used by the diagnostic standard MCS and the PM classification procedures. The six MCS–/PM+ disagreements affecting the estimate of the sensitivity of the PM to CAS (Participants 6–11), however, are concluded to be questionable because of methodological constraints.

Analyses of the Five MCS+/PM– Classification Disagreements

Demographics and speech competence variables. The first set of information in Table 3 provides age and sex data for the five participants classified CAS+ by MCS criteria (MCS+) but CAS– on PM criteria (PM–). The average age of the five MCS+/PM– participants (13.6 years; $SD = 7.8$) was not significantly older than the average age of participants (9.7 years; $SD = 4.8$) in the Consensus CAS+ group (effect size: 0.73; 95% CI [–0.23, 1.69]) but was significantly older than the average age of participants (7.6 years; $SD = 1.5$) in the Consensus CAS– group (effect size: 1.51; 95% CI [0.41, 2.61]). (All effect sizes are Hedges corrected; Hedges & Olkin, 1985.) It is possible that older participants such as Participants 1–5 are more likely to be classified as CAS+ by markers weighted by signs in the MCS that reflect persistent articulatory errors. As suggested previously, a potential strength of the PM is that inappropriate between-words pauses are essentially independent of developmental mastery of the segmental and suprasegmental domains of a language.

The sex distribution among the five participants classified MCS+/PM– (60% male, 40% female) was not significantly different from the distributions among participants in the Consensus CAS+ group (58.1% male, 41.9% female; Fisher exact p value = 1.000) or the Consensus CAS– group (75% male, 25% female; Fisher exact p value = .598). As with other findings to be discussed later in this supplement article, the relatively small number of participants in both CAS consensus groups precludes external generalizations. If replicated, however, the essentially equal proportion of male and female participants in the Consensus CAS+ group in the present sample is not consistent with estimates from convenience samples of children with idiopathic CAS

Table 3. Mayo Clinic System (MCS) and Pause Marker (PM) classification findings for childhood apraxia of speech (CAS).

Participants	CAS classification			Demographics		Speech competence		PM			Supplemental PM Signs (SPMS)			PM classification		
	Group	No.	MCS	PM	Age (years)	Sex	PCC (%)	PVC (%)	Pause opportunities	Inappropriate Type I pauses	PM score (%)	Rate	Stress	Transcoding	PM description	PM classification
Consensus groups																
CAS+ (n = 31)		+	+													
<i>M</i>				9.7	58.1% M/ 41.9% F	72.6	87.8	98.3	13.2	86.2						
<i>SD</i>				4.8		12.2	8.1	33.2	8.4	8.0						
Range				4–23		36.8–94.3	62.9–98.4	46–167	3–36	62.1–95.5						
% positive											75.0	78.6	78.6			
CAS– (n = 16)		–	–													
<i>M</i>				7.6	75.0% M/ 25.0% F	83.1	93.1	122.0	1.4	99.0						
<i>SD</i>				1.5		12.2	3.8	33.0	1.7	1.2						
Range				4–10		44.9–98.4	86.7–99.6	80–179	0–5	96.7–100.0						
% positive											33.3	18.8	46.7			
Classification disagreements																
MCS+/PM– (n = 5)		+	–													
1	1	+	–	5	M	74.6	96.4	120	0	100.0	0	0	0	PM–/SPMS–	PM–	
2	2	+	–	8	M	75.0	88.1	111	3	97.3	1	0	1	PM–/SPMS+	PM–	
3	3	+	–	14	M	89.9	98.5	144	5	96.5	1	0	1	PM–/SPMS+	PM–	
4	4	+	–	16	F	94.5	97.1	154	7	(95.5)	0	1	0	(PM)/SPMS–	(PM–)	
5	5	+	–	25	F	88.7	97.0	85	3	96.5	1	0	0	PM–/SPMS–	PM–	
<i>M</i>				13.6	60.0% M/ 40.0% F	84.5	95.4	122.8	3.6	97.2						
<i>SD</i>				7.8		9.1	4.2	27.4	2.6	1.7						
Range				5–25		74.6–94.5	88.1–98.5	85–154	0–7	95.5–100.0						
% positive											60.0	20.0	40.0			
MCS–/PM+ (n = 6)		–	+													
6	6	–	+	6	M	77.1	97.1	181	14	92.3	1	1	1	PM+/SPMS+	PM+	
7	7	–	+	8	F	54.5	90.5	50	7	86.0	1	0	1	PM+/SPMS+	PM+	
8	8	–	+	8	M	68.5	86.6	126	8	93.7	0	0	0	PM+/SPMS–	PM+	
9	9	–	+	11	M	70.8	89.1	97	6	93.8	*	1	1	PM+/SPMS+	PM+	
10	10	–	+	12	F	82.4	90.9	72	7	90.3	1	1	1	PM+/SPMS+	PM+	
11	11	–	+	17	F	77.8	90.7	113	18	84.1	1	1	0	PM+/SPMS+	PM+	
<i>M</i>				10.3	50.0% M/ 50% F	71.8	90.8	106.5	10.0	90.0						
<i>SD</i>				3.9		9.9	3.5	45.7	4.9	4.1						
Range				6–17		54.5–82.4	86.6–97.1	50–181	6–18	84.1–93.8						
% positive											80.0	66.7	66.7			

Note. Table entries in parenthesis are termed “marginal” classifications. Due to the participants marginal status, no positive or negative description is assigned. An asterisk denotes missing data. SPMS = supplemental PM signs; PCC = percentage of consonants correct; PVC = percentage of vowels correct.

indicating a higher prevalence of boys with CAS (as high as 90%; American Speech-Language-Hearing Association, 2007). As indicated previously in a review of the information in Table 1, whereas the CASI group of 41 participants was 68.3% male and 31.7% female, the CASN group of 19 participants was 47.4% male and 52.6% female. Genetic studies of CAS should be informative on neurodevelopmental implications of sex-linked or sex-influenced genomic pathways.

As indicated in Table 2 of PM I, MCS diagnostic criteria for CAS are based on 10 indices of speech and prosody competence and precision, whereas the PM diagnostic criterion for CAS is based solely on the frequency of occurrence of four types of inappropriate between-words pauses (resolution of marginal PM scores does use indices of segmental and suprasegmental competence and precision). A question is whether the MCS+/PM- classifications of the five participants in Table 3 might be associated with their lowered speech competence, to which only the MCS would be sensitive. The PCC data in Table 3 do not support this potential explanation for the five MCS+/PM- disagreements. The mean PCC for these five participants (84.5% [$SD = 9.1\%$]) was significantly higher than the mean PCC of the Consensus CAS+ participants (72.6% [$SD = 12.2\%$]; effect size: 0.98; 95% CI [0.01, 1.95]). None of the other three speech-competence comparisons, however, were statistically significant. To be specific, the MCS+/PM- participants' mean PCC score did not differ significantly from the mean PCC of participants in the Consensus CAS- group (83.1% [$SD = 12.2\%$]; effect size: 0.12; 95% CI [-0.89, 1.12]). Their percentage of vowels correct (PVC) also did not differ from the mean PVC of participants in the Consensus CAS+ group (effect size: 0.96; 95% CI [-0.01, 1.93]) or the Consensus CAS- group (effect size: 0.57; 95% CI [-0.45, 1.59]). Overall, the findings for the demographic and speech-competence variables for the five participants with MCS+/PM- classification outcomes do not suggest that differences in MCS versus PM classification outcomes were associated with their status on correlates of age, sex, or indices of consonant or vowel competence.

Speech-sampling variables. One sampling variable possibly associated with classification disagreements is the number of pause opportunities, with fewer pause opportunities (beyond the minimum required 40) possibly providing less reliable samples for the PM diagnostic criteria. The first column in the PM section of Table 3 shows the number of pause opportunities for participants in the two consensus CAS groups and for each of the five participants with MCS+/PM- classification outcomes. The difference in pause opportunities for the two consensus CAS groups is statistically significant (effect size: 0.70; 95% CI [0.08, 1.32]), with Consensus CAS- participants averaging more pause opportunities (122.0; $SD = 33.0$) than the Consensus CAS+ participants (98.3; $SD = 33.2$). Although the average number of pause opportunities for the five MCS+/PM- participants (122.8; $SD = 27.4$) was more similar to that of the Consensus CAS- participants than to that of the Consensus CAS+ participants, there was no statistically

significant difference between the number of pause opportunities of the MCS+/PM- group and either of the two consensus CAS groups. Thus, the PM- classifications for the five participants were not biased by significantly lower or higher denominators used in the computation of their PM. The number of pause opportunities in conversational speech is generally dictated by the language competence of speakers, with less voluble speakers requiring longer samples to meet type-token criteria for continuous speech samples (i.e., first-occurrence words; Shriberg & Kwiatkowski, 1983). Thus, the PM- findings for the five MCS+/PM- participants were not likely due to fewer opportunities for inappropriate pauses.

PM findings supporting PM- classifications. The prior discussions were focused on possible methodological constraints on the diagnostic accuracy of the five PM- classification outcomes for participants classified MCS+. The following findings review analyses interpreted as support for the PM- classification (i.e., CAS-) for each of the five participants.

The second and third columns in the PM section of Table 3 show the number of inappropriate Type I pauses and the PM score for each of the five MCS+/PM- speakers. The PM score data clearly classify the five participants as PM-. Four of the five participants' PM% scores met the criterion for PM- classification (i.e., were above the 96.0% cutoff point for classification as CAS-), with Participant 4's PM score of 95.5 meeting the criterion for a marginal PM score. There was a statistically significant difference between the five participants' mean PM score (97.2%; $SD = 1.7\%$) and the mean PM score of participants meeting requirements for the Consensus CAS+ group (86.2%; $SD = 8.0\%$; effect size: 1.43; 95% CI [0.43, 2.43]). On an individual basis, each of the scores of the five MCS+/PM- participants is more than 1 SD above the mean of the Consensus CAS+ group. We find it interesting that three of the five scores are less than 1 SD from the mean of the Consensus CAS- group, making their scores intermediate between CAS- and the cutoff score of 94% for CAS+.

The primary validity support for a PM- classification for the five MCS+/PM- participants is findings shown in the three columns in the section of Table 3 marked "SPMS." As described in PM I, SPMS criteria are used in two ways. First, for participants with marginal PM scores (i.e., 94.0%–95.9%), Rate, Stress, and Transcoding status is used to determine CAS status. Speakers in this marginal CAS range are classified as Marginal PM+ if they meet positive criteria for at least two of the three supplemental signs, Marginal PM- if they do not meet these criteria, and Indeterminate if missing SPMS data preclude resolution of the PM status. This summary of participants' PM and SPMS status is shown in the "PM classification" section called "PM description." The final column of this section gives the final PM classification. As shown in Table 3, Participant 4's PM classification of marginal PM- was due to the SPMS- outcome.

The second purpose of the SPMS is to use information on these three signs to support the validity of classification decisions on the basis of the PM. Notice that 75.0%,

78.6%, and 78.6% of the Consensus CAS+ participants were positive for deficits in, respectively, Rate, Stress, and Transcoding. Among the five MCS+/PM- participants, by contrast, 60%, 20%, and 40% were coded positive for those deficits. These percentage findings are closer to those for the Consensus CAS- participants, with 33.3%, 18.8%, and 46.7% meeting criteria. Using the minimum requirement of two of the three SPMS for a positive SPMS outcome, the SPMS data do not support CAS+ classification for three of the five MCS+/PM- participants.

Summary. The present PM sensitivity finding of 86.8% is considered strong but does not reach conclusive support (at least 90%). The findings reviewed in this section do not support a demographic, speech-severity, or PM sampling variable as a possible source of methodological or sampling bias for the five MCS+/PM- classification outcomes that attenuate the estimated sensitivity of the PM. Analyses of SPMS data suggest that the findings for at least some of the five MCS+/PM- participants are more similar to those for the Consensus CAS- participants than the Consensus CAS+ participants. Without a methodological rationale for viewing the five MCS+/PM- classification outcomes as questionable, however, the present sensitivity of the PM to true positive CAS using MCS classifications as the validation standard is estimated to be 86.8% (see Figure 1).

The Six MCS-/PM+ Classification Disagreements

The lowest section of Table 3 gives information for six participants who are suspected to be positive for CAS and were classified as negative for CAS using the MCS but positive using the PM. As shown in Figure 1C, these six MCS-/PM+ classification outcomes for the combined CASI and CASN groups attenuated the PM specificity estimate to 72.7%. The following examination of the six disagreements is analytically similar to that just conducted for the five MCS+/PM- classification disagreements. Findings from this examination, however, are interpreted as support for viewing the specificity estimate on the basis of the CAS group findings as questionable because of methodological constraints. As shown in Figure 1, these classification outcomes are excluded from the reported estimate of the specificity of the PM.

Demographic, speech competence, and PM sampling variables. As shown in Table 3, the average age of the six MCS-/PM+ participants (10.3 years; $SD = 3.9$) was significantly older than the average age of the Consensus CAS- participants (7.6 years; $SD = 1.5$; effect size: 1.11; 95% CI [0.11, 2.10]). There were no other significant findings for the sex percentages, average speech competence (PCC, PVC), or average PM pause opportunities of the six MCS-/PM+ participants compared with the Consensus CAS+ and Consensus CAS- participants. Thus, other than the finding that the MCS-/PM+ participants were significantly older than the Consensus CAS- participants, there were no significant differences in the demographic data, speech competencies, or opportunities for inappropriate pauses in the MCS-/PM+ participants that may have influenced their classification outcomes.

Missing data constraints. There are two variables that may have played a significant role in the MCS- classifications for each of the six MCS-/PM+ participants. First, as part of the MCS classification procedures, the second author had provided, for each of the participants who was suspected by the referral source to be positive for CAS, a spreadsheet containing anecdotal comments on the classification outcomes. Examination of these comments indicated that missing data on MSAP tasks were a significant concern for the MCS classification decisions for Participants 6–8. Specifically cited were the lack of participant scores on one to three of the following five MSAP tasks (see Table 2): Diadochokinesis Task, Nonword Repetition Task, Syllable Repetition Task, Vowel Task 1, and Vowel Task 2. Participant responses to these tasks presumably are important for classifying or diagnosing CAS using the criterion number of MCS signs on the minimal number of speech tasks required, as listed in Table 2 of PM I. Thus, because of a concern that missing data on one or more of these tasks may have constrained MCS classification outcomes for these three participants, it was deemed appropriate to view the MCS-/PM+ classification disagreements for these participants as questionable.

A second methodological constraint identified in the anecdotal comments is a terminological issue that affected MCS classification outcomes for Participants 9–11. These three participants are siblings whose chromosome translocations and histories of extensive speech services for CAS have been described in a clinical report (Shriberg, Jakielski, & El-Shanti, 2008). Although none of the three participants met MCS criteria for CAS, summaries of speech findings in the spreadsheet included the following anecdotal comments for Participants 9, 10, and 11, respectively: “There is the percept of mild motor impairment associated with cognitive impairment,” “may have mild motor speech impairment (possible),” and “there are minor characteristics of motor speech disorder.” There were no anecdotal comments for these participants indicating that the motor speech disorder was consistent with dysarthria or further specifying the type of motor speech disorder. Although none of the three participants met MCS criteria for CAS+, these anecdotal comments on motor speech involvement are viewed as support for also viewing these classification disagreements as questionable.

PM variable findings supporting PM+ classifications. Support for classification of the six MCS-/PM+ participants as CAS+ can be marshaled from two additional sources of information. The primary source of information for this perspective is the SPMS data in Table 3. As shown in the bottom row, 80.0%, 66.7%, and 66.7% of the MCS-/PM+ participants met criteria for CAS on, respectively, Rate, Stress, and Transcoding. Five out of the six participants (83.3%) met the SPMS criterion for CAS+ (positive findings for least two of the three signs). These percentages are more consistent with the percentage data in Table 3 for the Consensus CAS+ participants than for the Consensus CAS- participants.

The second source of support for classifying these participants as CAS+ is discussed and illustrated fully in

the PM technical report (Tilkens et al., 2017), which provides descriptive statistical findings for the eight subtypes of inappropriate pauses described in PM I but not included in Table 3. In essence, inappropriate pauses that meet the criteria described in Table 7 of PM I for *abrupt* are the most frequently occurring of the four Type I inappropriate pauses and are theoretically most coherent with speech processing deficits in apraxia of speech (see PM III). Additional support for classifying the present six participants as CAS+ can be marshaled by comparing their proportional occurrence of this type of inappropriate pause to the statistically lower proportion of abrupt inappropriate pauses in Consensus CAS- participants.

Summary. Anecdotal information on MCS classifications, findings reported in Table 3, and other findings are proposed as support for viewing each of the six MCS-/PM+ classification disagreements described in Table 3 as questionable. As shown in Figure 1C, excluding these classification outcomes (i.e., disagreements) from the calculation yields a PM specificity estimate of 100% for children and adults with suspected CAS.

AAS Analyses

As reviewed previously, the consensus in both the developmental and adult-onset apraxia of speech literatures is that the unique neurocognitive and neuromotor signature of CAS occurs primarily in the process of transcoding linguistic representations into the plans and programs for articulate speech. As indicated previously, there is also increasing empirical support for core CAS deficits in representational processes and in feed-forward elements of transcoding. In the present context, therefore, findings supporting the ability of the PM to identify AAS would support both the construct validity and psychometric accuracy of the PM. The following analyses examined the sensitivity of the PM to two forms of AAS: AAS due to a neurological incident (AOS) and AAS due to neurodegenerative disease (PPAOS). The analyses are currently limited to PM sensitivity questions because specificity estimates for AAS require a sufficiently large sample of adults with representative subtypes of dysarthria, documented using quantitative perceptual and acoustic indices.

MCS/PM Classification Agreements

As described earlier, the second author, using an adult form of the MCS, classified 31 participants with adult-onset speech disorder as positive for either AOS or PPAOS. Figure 2 gives the percentages of agreement between these classifications and classifications using the PM procedures. As shown, MCS+/PM+ classification agreements occurred for 10 of the 15 (66.7%) participants with AOS in Figure 2A and 12 of the 16 (75%) participants with PPAOS in Figure 2B, yielding a total sensitivity estimate for the PM for AAS of 71.0%. As shown in Figure 2C, with the remaining nine speakers (the “Voicers”) removed from the computation, the PM was 100% sensitive to 22 speakers with AOS or PPAOS forms of AAS classified using MCS criteria.

Analyses of the Nine MCS+/PM- Classification Disagreements

Table 4 gives summary information for the two groups of participants with AAS, organized in the same format as Table 3. The top section provides summary statistics for the 22 Consensus AAS+ participants. As shown in the PM column called “PM score,” they averaged approximately the same PM score (87.0%; $SD = 7.3\%$) as the 31 Consensus CAS+ participants in Table 3 (PM = 86.2%; $SD = 8.0\%$). The other two sections in Table 4 give individual data for the nine MCS+/PM- participants termed Voicers, including five participants with AOS (Participants 1–5) and four with PPAOS (Participants 6–9). The following discussions examine methodological and substantive variables that might inform PM findings as support for the construct validity and sensitivity of the PM to identify speakers with AAS.

Demographic and speech competence variables. There were no significant differences between the average age of participants in the Consensus AAS+ (MCS+/PM+) group (67.3 years; $SD = 11.8$) and those of the five participants in the AOS Voicers group classified as MCS+/PM- (60.4 years; $SD = 9.7$; effect size: -0.58 ; 95% CI $[-1.57, 0.40]$) or the four participants in the PPAOS Voicers group classified as MCS+/PM- (71.8 years; $SD = 5.6$; effect size: 0.39 ; 95% CI $[-0.68, 1.46]$). There were also no significant differences between the sex proportions for participants in the Consensus AAS+ group (59.1% male, 40.9% female) and those for participants in the MCS+/PM- AOS Voicers group (80.0%; Fisher exact p value = .621) or speakers in the MCS+/PM- PPAOS Voicers group (75.0%; Fisher exact p value = .639). Last, there were no significant differences between the speech competence of participants in the Consensus AAS+ group (PCC: $M = 90.7\%$; $SD = 7.1\%$; PVC: $M = 94.0\%$; $SD = 5.8\%$) and those of participants in the MCS+/PM- AOS Voicers group (PCC: $M = 96.2\%$; $SD = 3.5\%$; effect size: 0.80 ; 95% CI $[-0.19, 1.80]$; PVC: $M = 98.4\%$; $SD = 2.5\%$; effect size: 0.79 ; 95% CI $[-0.20, 1.78]$) or participants in the MCS+/PM- PPAOS Voicers group (PCC: $M = 96.4\%$; $SD = 1.6\%$; effect size: 0.83 ; 95% CI $[-0.26, 1.92]$; PVC: $M = 97.3\%$; $SD = 0.80$; effect size: 0.59 ; 95% CI $[-0.49, 1.67]$).

PM variables. As shown in the PM column in Table 4 called “Pause opportunities,” there was considerable variability in the number of pause opportunities in the speech samples of the three groups of adult participants with AAS (Consensus AAS+, AOS Voicers, and PPAOS Voicers). However, there were no statistical differences between the average number of pause opportunities for the Consensus AAS+ participants ($M = 120.0$; $SD = 45.5$) and those for the MCS+/PM- AOS Voicers ($M = 155.8$; $SD = 33.8$; effect size: 0.79 ; 95% CI $[-0.20, 1.79]$) or the MCS+/PM- PPAOS Voicers ($M = 174.0$; $SD = 119.8$; effect size: 0.87 ; 95% CI $[-0.22, 1.96]$).

SPMS findings. Table 4 also includes information on the SPMS criterion findings for the Consensus AAS+ group and for the nine participants in the AOS and PPAOS Voicer groups who did not meet the PM score for AOS.

Figure 2. Sensitivity estimates for the Pause Marker based on the Mayo Clinic System and Pause Marker classification outcomes for the 31 participants with adult-onset apraxia of speech, for participants in (A) the group with apraxia of speech ($n = 15$), (B) the group with primary progressive apraxia of speech ($n = 16$), and (C) the two groups combined ($n = 31$).

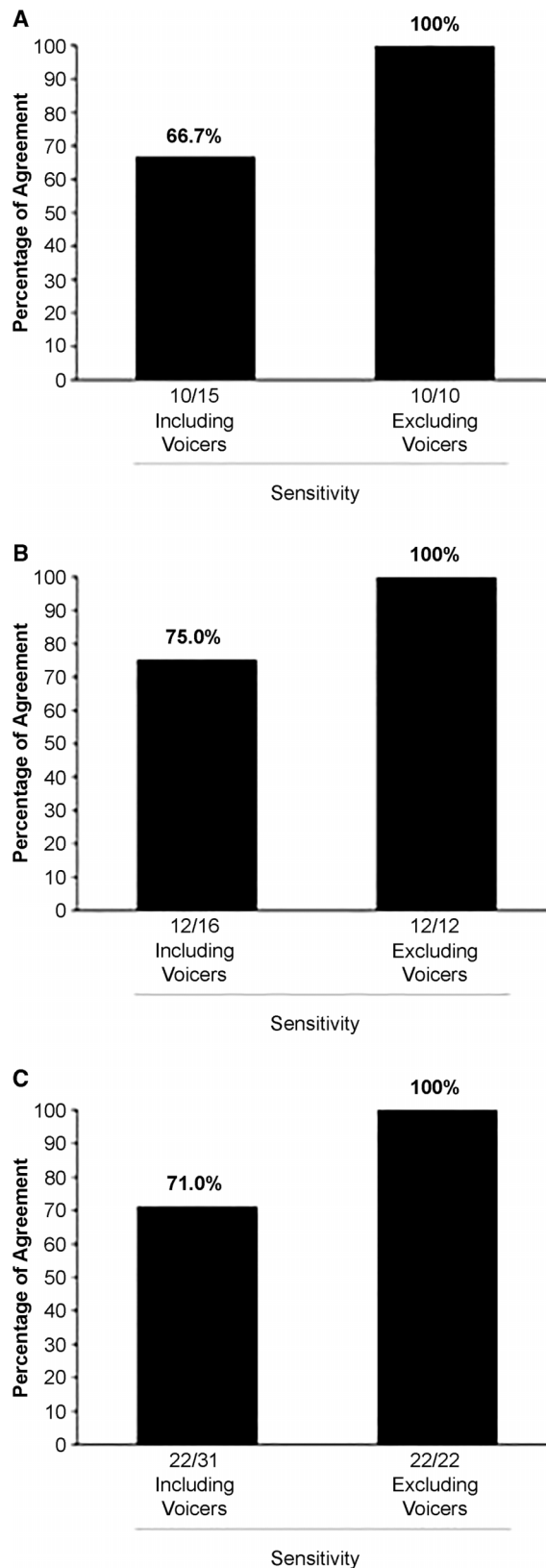


Table 4. Mayo Clinic System (MCS) and Pause Marker (PM) classification findings for adult-onset apraxia of speech (AAS).

Participants	CAS classification		Demographics		Speech competence		PM			Supplemental Pause Marker Signs (SPMS)			PM classification			
	Group	No.	MCS	PM	Age (years)	Sex	PCC (%)	PVC (%)	Pause opportunities	Inappropriate Type I pauses	PM score (%)	Rate	Stress	Transcoding	PM description	PM classification
Consensus AAS+ (n = 22)			+	+												
M					67.3	59.1% M/ 40.9% F	90.7	94.0	120.0	14.2	87.0					
SD					11.8		7.1	5.8	45.5	8.4	7.3					
Range					45–84		68.9–97.6	76.9–99.5	45–202	4–40	73.1–95.5					
% positive												95.5	86.4	50.0		
AOS Voicers (n = 5)			+	–												
1	1		+	–	68	F	99.4	100.0	149	1	99.3	0	0	*	PM–/SPMS–	PM–
2	2		+	–	55	M	99.4	99.6	139	5	96.4	0	0	*	PM–/SPMS–	PM–
3	3		+	–	50	M	97.1	100.0	115	2	98.3	1	0	*	PM–/SPMSI	PM–
4	4		+	–	56	M	92.1	94.0	204	0	100.0	0	0	*	PM–/SPMS–	PM–
5	5		+	–	73	M	93.1	98.4	172	9	(94.8)	1	0	*	(PM)/SPMSI	(PM–)
M					60.4	80.0% M/ 20.0% F	96.2	98.4	155.8	3.4	97.8					
SD					9.7		3.5	2.5	33.8	3.6	2.1					
Range					50–73		92.1–99.4	94.0–100.0	115–204	0–9	94.8–100.0					
% positive												40.0	0.0	*		
PPAOS Voicers (n = 4)			+	–												
1	1		+	–	67	F	97.4	98.2	101	2	98.0	1	1	0	PM–/SPMS+	PM–
2	2		+	–	78	M	95.8	96.8	100	2	98.0	1	1	1	PM–/SPMS+	PM–
3	3		+	–	68	M	94.5	96.6	351	7	98.0	1	0	1	PM–/SPMS+	PM–
4	4		+	–	75	M	97.9	97.8	144	1	99.3	1	1	1	PM–/SPMS+	PM–
M					71.8	75.0% M/ 25.0% F	96.4	97.3	174.0	3.0	98.3					
SD					5.6		1.6	0.8	119.8	2.7	0.6					
Range					67–78		94.5–97.9	96.6–98.2	100–351	1–7	98.0–99.3					
% positive												100.0	75.0	75.0		

Note. Table entries in parenthesis are termed “marginal” classifications. Due to the participants marginal status, no positive or negative description is assigned. An asterisk denotes missing data. SPMS = supplemental PM signs; PCC = percentage of consonants correct; PVC = percentage of vowels correct; AOS = apraxia of speech; SPMSI = indeterminate SPMS; PPAOS = primary progressive apraxia of speech.

After we adjusted for missing data and used SPMS findings as needed to resolve marginal PM scores, the SPMS findings supported the MCS+ classifications of participants in two of the three groups. Of the 22 participants in the Consensus AAS+ group, 95.5%, 86.4%, and 50.0% had inappropriate Rate, Stress, and Transcoding scores, respectively. Transcoding scores were not available for any of the five AOS Voicer participants. A total of 40.0% had inappropriate Rate scores, and none had inappropriate Stress scores. Moreover, of the three participants for whom an SPMS score could be determined (Participants 1, 2, and 4), each was classified as PM-. Last, for the four PPAOS Voicers, 100.0%, 75.0%, and 75.0% had inappropriate Rate, Stress, and Transcoding scores, respectively. As shown in the "PM description" column, none of the four participants met the PM criteria for apraxia of speech, but all four met the SPMS criteria for it.

Voicers summary. The findings summarized in Figure 2 and Table 4 indicate that the PM was sensitive to AAS for 71.0% of participants with AAS, but insensitive to it in the remaining 29.0% of the participants in the present sample of speakers with AOS or PPAOS. Examination of demographic, speech competence, and pause-sampling data for the latter participants did not identify any sampling constraints that might account for their PM- classifications. Acoustic analyses of their recordings, however, indicated that each of the nine speakers had a speaking style of near-continuous voicing. There is no case-history information on whether this speaking style was present before the onset of AOS or PPAOS. This behavior was not observed in speakers of any age among the 60 participants who were suspected to be positive for CAS or in any of the participants with SD (discussed in the next section). If the onset of this speaking style was associated with AAS, it might have been developed to avoid inappropriate pauses, much as typical speakers use pause fillers and other behaviors to avoid pauses.

Summary. PM scores using the same cutoff criteria to identify CAS were obtained from 31 participants with AAS as identified using MCS adult AOS criteria. Our rationale for including speakers with AAS was to determine whether findings support the construct validity of the PM with participants who were suspected to be positive for CAS and with participants with SD (discussed in the next section). Findings for 71.0% of the AAS participants support the construct validity of the PM, but the PM yielded false negatives for a group of speakers whose near-continuous voicing is incompatible with inappropriate pauses.

SD Analyses

The two previous analyses and examinations of findings for MCS/PM disagreements have assessed the diagnostic accuracy of the PM in participants who were suspected to be positive for CAS in idiopathic and neurogenetic contexts and construct validity support for the PM in participants with two types of AAS. The primary focus in the present findings, and the focus of this research, is the

specificity of the PM to discriminate CAS from SD. As described earlier, none of the 205 participants with SD (ages 3–9 years) were also suspected to be positive for CAS. To be specific, each participant met the SDCS criterion for an idiopathic speech sound disorder characterized by age-inappropriate phoneme deletions and/or substitutions in conversational speech, as quantified and classified by software in Programs to Examine Phonetic and Phonological Evaluation Records (Shriberg et al., 2001). As shown in Figure 1 of PM I and discussed previously, for the following analyses it is not necessary to subtype participants with idiopathic SD into the three putative causal pathways for their speech delays (i.e., neurolinguistic and psycholinguistic constraints associated with genomic variables, otitis media with effusion, and psychosocial variables).

Consensus SD Group

Figure 3 and Table 5 provide summary PM specificity findings for the four cohorts of participants with SD described in Table 1 (SD1–SD4) and for the four cohorts combined. As shown in Figure 3, specificity estimates for the four cohorts ranged from 90.9% (SD2) to 100% (SD3 and SD4), averaging 98.5% across the 205 participants. As shown at the bottom of the figure, PM+ classification findings were obtained for two participants in SD2 and one in SD1. The following review of information in Table 5 is relevant to whether or not the PM scores for these three participants (1.5% of the 205 participants with SD) should be viewed as false positive CAS classification outcomes.

Analyses of the Three SDCS/PM Disagreements

Demographic and speech competence variables. Compared with the average age (4.4 years; $SD = 1.1$) and sex distribution (70.8% male, 29.2% female) of the 202 SD

Figure 3. Specificity estimates for the Pause Marker on the basis of Pause Marker classification outcomes for the 205 participants with speech delay (SD) in the four SD cohorts and for the combined SD groups.

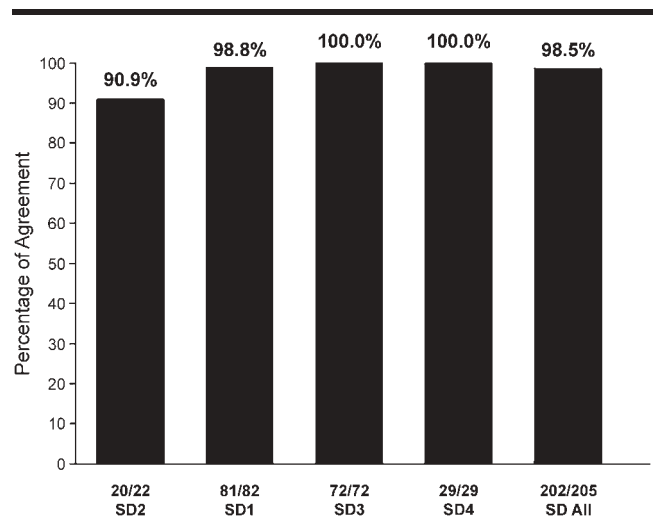


Table 5. Specificity information for the three participants with speech delay who were classified positive by the Pause Marker (PM+) compared with 202 participants classified negative by the PM (PM-).

Participants	CAS classification		Demographics		Speech competence		PM			SPMS			PM classification			
	Group	No.	SDCS	PM	Age (years)	Sex	PCC (%)	PVC (%)	Pause opportunities	Inappropriate Type I pauses	PM score (%)	Rate	Stress	Transcoding	PM description	PM classification
PM-																
(<i>n</i> = 202)		-	-													
<i>M</i>					4.4	70.8% M/ 29.2% F	72.3	92.6	96.3	1.0	99.0					
<i>SD</i>					1.1		11.5	5.1	32.0	1.3	1.3					
Range					3–9		17.5–99.1	70.1–100.0	40.0–195.0	0.0–6.0	94.4–100.0					
% positive											23.9	16.9	34.7			
PM+																
(<i>n</i> = 3)		-		+												
1		-		+	3	M	89.8	95.8	63	5	92.1	+	+	*	PM+/SPMS+	PM+
2		-		+	5	M	81.9	95.3	48	6	87.5	-	-	+	PM+/SPMS-	PM+
3		-		+	6	F	87.0	95.1	137	6	(95.6)	+	-	+	(PM)/SPMS+	(PM+)
<i>M</i>					4.7	66.7% M/ 33.3% F	86.3	95.4	82.7	5.7	91.7					
<i>SD</i>					1.5		4.0	0.3	47.6	0.6	4.1					
Range					3–6		81.9–89.8	95.1–95.8	48–137	5–6	87.5–95.6					
% positive											66.7	33.3	100.0			

Note. Table entries in parenthesis are termed “marginal” classifications. Due to the participants marginal status, no positive or negative description is assigned. An asterisk denotes missing data. CAS = childhood apraxia of speech; SPMS = supplemental PM signs; SDCS = Speech Disorders Classification System; PCC = percentage of consonants correct; PVC = percentage of vowels correct.

participants with PM– scores, Table 5 shows that there were no significant differences in the average ages of the three SD participants with PM+ scores (4.7 years; $SD = 1.5$; effect size: 0.27; 95% CI [–0.87, 1.41]) or in sex distribution (66.7% male, 33.3% female; Fisher exact p value = 1.000). Compared with the average PCC (72.3%; $SD = 11.5\%$) and PVC (92.6%; $SD = 5.1\%$) scores of the PM– group, there were no significant differences in the PCC (86.3%; $SD = 4.0$; effect size: 1.22; 95% CI [0.07, 2.36]) or PVC (95.4%; $SD = 0.3$; effect size: 0.55; 95% CI [–0.59, 1.69]) scores of the three PM+ SD participants. Thus, age, sex, and severity of speech involvement were not associated with the PM+ classification outcomes for these three participants, each of whom met SDCS criteria for SD (age-inappropriate speech sound deletions and/or substitutions in continuous speech) at assessment.

PM variables. As described previously, the number of opportunities for inappropriate pauses is viewed as indexing language productivity. As shown in the first column in the PM section in Table 5, the 202 participants with SD who had PM– scores averaged 96.3 ($SD = 32.0$) pause opportunities. The three SD participants with PM+ classification outcomes averaged 82.7 ($SD = 47.6$) pause opportunities, a nonsignificant difference (effect size: –0.42; 95% CI [–1.56, 0.72]).

SPMS and PM classification outcomes. Although the three participants' classification outcomes were PM+ (see Table 5, rightmost columns), they each had different "PM description" entries (second column from the right). Participant 1 had positive (+) outcomes on both the SPMS criterion and the PM score, Participant 2 did not reach criterion on the SPMS, and Participant 3 had a marginally positive PM score and a positive outcome on the SPMS.

Summary. The overall specificity findings for the PM of 98.5% are based on results indicating that three of 205 participants who met criteria for SD in their original study samples were positive for CAS when classified using the PM. Close examination of the case histories of these participants would be informative in attempting to determine whether their clinical diagnoses and SDCS classifications as having SD in their original study were false negatives for CAS or their PM classifications as CAS+ in the present study are false positives. This initial estimate of the specificity of the PM to discriminate CAS from SD requires cross-validation by other research groups using participants with diverse demographic, language, speech, and complex neurodevelopmental backgrounds. As discussed in PM I, there is also a need to assess the specificity of the PM in well-defined groups of children and adolescents with subtypes of idiopathic dysarthria in idiopathic contexts and in complex neurodevelopmental disorders.

Summary and Conclusions

We have identified three principal findings and conclusions from the research described in this supplement article:

1. The PM identified as positive for CAS 86.8% of the participants classified by MCS criteria as positive for

CAS. Our discussion considered variables plausibly associated with the five classification disagreements that attenuated this estimate of the sensitivity of the PM to CAS. Adjusting for classification outcomes ruled questionable because of sampling constraints, the PM also identified as negative for CAS 100% of the speakers who were classified by MCS criteria as not having CAS. Within the sampling constraints of this study (i.e., demographic and other possible limitations on representativeness), the diagnostic accuracy data (likelihood ratios) for this sample of speakers are interpreted as near conclusive, contingent on the validity of the rationale for and interpretation of the disagreements that attenuated the sensitivity and specificity estimates.

2. For a group of speakers identified as having AAS by an adult version of the MCS used to assess construct validity support for the PM to identify people with apraxia of speech, the PM correctly identified 71% of the speakers as positive for AAS. A constraint on this initial sensitivity estimate was that nine of the participants with AAS, termed Voicers, had a phonatory style that precluded inappropriate pauses. With these speakers excluded, PM sensitivity was 100%. The continuous or near-continuous voicing observed in these nine speakers was not observed in any of the other child or adult participants assessed using PM methods. It was not possible to obtain information on these speakers' premorbid speech. Recall that the AAS group was included only for construct validity questions (see PM III); the PM was not developed for clinical-research use with persons with AAS. Additional studies assessing the prevalence and topography of continuous voicing in persons with AAS could be theoretically and clinically informative.
3. Using the SDCS as the reference standard to identify SD, the PM correctly classified 98.5% of the participants with SD as not meeting PM criteria for CAS. Our discussion considered variables plausibly associated with the three classification disagreements attenuating this estimate of the specificity of the PM, the primary question posed in this research.

The results of the present study are interpreted as supporting the PM as a psychometrically robust diagnostic marker of CAS that meets validity criteria proposed in Table I of PM I to be the most important attribute of highly valued diagnostic markers. Two following articles address, in turn, the theoretical coherence of the PM with the core speech processing deficits posited for speakers with CAS and the generality of the PM to scale severity of CAS.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grant DC000496,

awarded to Lawrence D. Shriberg, and by a core grant, National Institute of Child Health and Development Grant HD03352, to the Waisman Center. The sixth and ninth authors made the original and substantial continuing contributions that led to the development of the Pause Marker. We are grateful to the following colleagues and collaborators for their significant contributions to this research: Len Abbeduto, Nancy Alarcon, Becky Baas, Adriane Baylis, Richard Boada, Roger Brown, Stephen Camarata, Thomas Campbell, Richard Folsom, Lisa Freebairn, Jordan Green, Barbara Lewis, Christopher Moore, Katherine Odell, Bruce Pennington, Nancy Potter, Jonathan Preston, Erin Redle, Heather Leavy Rusiewicz, Alison Scheer-Cohen, Kristie Spencer, Ruth Stoeckel, Bruce Tomblin, Jennifer Vannest, and Emily White. We also thank the many participants, parents of participants, and research colleagues who have contributed insights into the needs and issues in diagnostic research in childhood apraxia of speech.

References

- American Speech-Language-Hearing Association.** (2007). *Childhood apraxia of speech* [Technical report]. Retrieved from <http://www.asha.org/public/speech/disorders/ChildhoodApraxia/>
- Dollaghan, C. A.** (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore, MD: Brookes.
- Duffy, J. R., & Josephs, K. A.** (2012). The diagnosis and understanding of apraxia of speech: Why including neurodegenerative etiologies may be important. *Journal of Speech, Language, and Hearing Research, 55*, S1518–S1522.
- Duffy, J. R., Strand, E. A., Clark, H., Machulda, M., Whitwell, J. L., & Josephs, K. A.** (2015). Primary progressive apraxia of speech: Clinical features and acoustic and neurologic correlates. *American Journal of Speech-Language Pathology, 24*, 88–100.
- Hedges, L. V., & Olkin, I.** (1985). *Statistical methods for meta-analysis*. Boston, MA: Academic Press.
- Josephs, K. A., Duffy, J. R., Strand, E. A., Machulda, M. M., Senjem, M. L., Master, A. V., . . . Whitwell, J. L.** (2012). Characterizing a neurodegenerative syndrome: Primary progressive apraxia of speech. *Brain, 135*, 1522–1536.
- Josephs, K. A., Duffy, J. R., Strand, E. A., Whitwell, J. L., Layton, K. F., Parisi, J. E., . . . Petersen, R. C.** (2006). Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. *Brain, 129*, 1385–1398.
- 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.** (2010). Childhood speech sound disorders: From postbehaviorism to the postgenomic era. In R. Paul & P. Flipsen, Jr. (Eds.), *Speech sound disorders in children* (pp. 1–33). San Diego, CA: Plural.
- Shriberg, L. D., Allen, C. T., McSweeney, J. L., & Wilson, D. L.** (2001). PEPPER: Programs to Examine Phonetic and Phonologic Evaluation Records [Computer software]. Madison: University of Wisconsin–Madison, Waisman Center, Research Computing Facility.
- Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeney, J. L., . . . Wilson, D. L.** (2010a). Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics, 24*, 795–824.
- Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeney, J. L., . . . Wilson, D. L.** (2010b). Perceptual and acoustic reliability estimates for the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics, 24*, 825–846.
- Shriberg, L. D., Jakielski, K. J., & El-Shanti, H.** (2008). Break-point localization using array-CGH in three siblings with an unbalanced 4q;16q translocation and childhood apraxia of speech (CAS). *American Journal of Medical Genetics: Part A, 146A*, 2227–2233.
- Shriberg, L. D., & Kwiatkowski, J.** (1983). Computer-assisted Natural Process Analysis (NPA): Recent issues and data. *Seminars in Speech and Language, 4*, 389–406.
- Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C.** (1990). *The Prosody-Voice Screening Profile*. Tucson, AZ: Communication Skill Builders.
- Shriberg, L. D., & Lohmeier, H. L.** (2008). *The Syllable Repetition Task* [Technical report]. Madison: University of Wisconsin–Madison, Waisman Center, Phonology Project.
- Shriberg, L. D., Lohmeier, H. L., Campbell, T. F., Dollaghan, C. A., Green, J. R., & Moore, C. A.** (2009). A nonword repetition task for speakers with misarticulations: The Syllable Repetition Task (SRT). *Journal of Speech, Language, and Hearing Research, 52*, 1189–1212.
- Shriberg, L. D., Potter, N. L., & Strand, E. A.** (2011). Prevalence and phenotype of childhood apraxia of speech in youth with galactosemia. *Journal of Speech, Language, and Hearing Research, 54*, 487–519.
- Shriberg, L. D., Strand, E. A., Fourakis, M., Jakielski, K. J., Hall, S. D., Karlsson, H. B., . . . Wilson, D. L.** (2017a). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: I. Development and description of the Pause Marker. *Journal of Speech, Language, and Hearing Research, 60*, S1096–S1117. https://doi.org/10.1044/2016_JSLHR-S-15-0296
- Shriberg, L. D., Strand, E. A., Fourakis, M., Jakielski, K. J., Hall, S. D., Karlsson, H. B., . . . Wilson, D. L.** (2017b). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: III. Theoretical coherence of the Pause Marker with speech processing deficits in Childhood Apraxia of Speech. *Journal of Speech, Language, and Hearing Research, 60*, S1135–S1152. https://doi.org/10.1044/2016_JSLHR-S-15-0298
- Shriberg, L. D., Strand, E. A., Fourakis, M., Jakielski, K. J., Hall, S. D., Karlsson, H. B., . . . Wilson, D. L.** (2017c). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: IV. The Pause Marker Index. *Journal of Speech, Language, and Hearing Research, 60*, S1153–S1169. https://doi.org/10.1044/2016_JSLHR-S-16-0149
- Shriberg, L. D., Strand, E. A., & Mabbie, H. L.** (2017). *Prevalence of speech and motor speech disorders in idiopathic speech delay and complex neurodevelopmental disorders*. Manuscript submitted for publication.
- Tilkens, C. M., Karlsson, H. B., Fourakis, M., Hall, S. D., Mabbie, H. L., McSweeney, J. L., . . . Shriberg, L. D.** (2017). *A diagnostic marker to discriminate childhood apraxia of speech (CAS) from Speech Delay (SD)*. (Technical Report No. 22). Retrieved from Phonology Project website: <http://www.waisman.wisc.edu/phonology/>
- Walter, S. D., & Cook, R. J.** (1991). A comparison of several point estimators of the odds ratio in a single 2 × 2 contingency table. *Biometrics, 47*, 795–811.