Research Article

The Apraxia of Speech Rating Scale: Reliability, Validity, and Utility

Joseph R. Duffy,[a] Peter R. Martin,[b] Heather M. Clark,[a] Rene L. Utianski,[a] Edythe A. Strand,[a,c] Jennifer L. Whitwell,[d] and Keith A. Josephs[a]

[a] Department of Neurology, Mayo Clinic, Rochester, MN [b] Department of Quantitative Health Sciences (Biostatistics), Mayo Clinic, Rochester, MN [c] Department of Speech & Hearing Sciences, University of Washington, Seattle [d] Department of Radiology, Mayo Clinic, Rochester, MN

ABSTRACT

Purpose: The purpose of this study was to examine the interrater reliability and validity of the Apraxia of Speech Rating Scale (ASRS-3.5) as an index of the presence and severity of apraxia of speech (AOS) and the prominence of several of its important features.
Method: Interrater reliability was assessed for 27 participants. Validity was examined in a cohort of 308 participants (120 with and 188 without progressive AOS) through item analysis; item-Total score correlations; correlations among ASRS Total score and component subscores and independent clinical ratings of AOS, dysarthria and aphasia severity, intelligibility, and articulatory errors, as well as years postonset and age; and regression models assessing item and Total score prediction of AOS presence.
Results: Interrater reliability was good or excellent for most items and excellent for the Total score. Item and Total score analyses revealed good separation of participants with versus without AOS. Inter-item and item-Total score correlations were generally moderately high as were correlations between the ASRS Total score and independent ratings of AOS severity, intelligibility, and articulatory errors. The Total score was not meaningfully correlated with ratings of aphasia and dysarthria severity, years postonset, or age. Total scores below 7 and above 10 revealed excellent diagnostic sensitivity and specificity for AOS. The presence of eight or more abnormal features was also highly predictive of AOS presence.
Conclusions: The ASRS-3.5 is a reliable and valid scale for identifying the presence and severity of AOS and its predominant features. It has excellent sensitivity to AOS presence and excellent specificity relative to aphasia and dysarthria in patients with neurodegenerative disease.
Supplemental Material: https://doi.org/10.23641/asha.21817584

The assessment, description, and diagnosis of apraxia of speech (AOS) have been a longstanding challenge to clinicians and researchers (e.g., Ballard et al., 2016; Haley et al., 2012; McNeil et al., 2017; Wambaugh et al., 2019). The reasons for this include, at the least, (a) evolving concepts about the nature of AOS and its core clinical features, (b) overlap of some features of AOS with those associated with frequently co-occurring aphasia and dysarthria, (c) challenges to the reliability of perceptually based diagnostic judgments, and (d) the incomplete development of formal metrics for describing the disorder.

There are important clinical reasons for using perceptual measures as the gold standard for diagnosis (McNeil et al., 2017), but the need persists for a quantifiable measure of the presence and severity of AOS that is separable from aphasia and dysarthria. The Apraxia of Speech Rating Scale (ASRS; Josephs et al., 2012; Strand et al., 2014) has shown promise in this regard. Its further development, with emphasis on its reliability, validity, and data relevant to its clinical use, was the focus of this study.

Background

Recognizing the lack of a widely accepted measure for describing AOS, an initial version of the ASRS was developed during research requiring a clinical description

Correspondence to Joseph R. Duffy: jduffy@mayo.edu. *Disclosure: The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

of progressive AOS (PAOS) and examination of its relationship with other clinical findings, neuroimaging, and molecular pathology; for a recent summary, see Duffy et al. (2021) or Josephs et al. (2021). In 2012, a study characterizing primary progressive apraxia of speech (PPAOS) used a 16-item "AOS rating scale," hereafter referred to as the ASRS-1, to describe the salient features of the disorder (Josephs et al., 2012). The ASRS-1 and its subsequent modifications have been used by our and other research groups in studies in which it has been important to describe and quantify AOS; it thus initially evolved in service to broader research goals. Only after its utility in several studies became apparent did we recognize its potential for broader application. Consequently, the steps to establish its validity and reliability have not followed the usual process associated with rating scale development.

Validity and Uses of the ASRS

The ASRS-1 has distinguished groups with PAOS from neurotypical speakers (Duffy et al., 2017) and groups with left hemisphere stroke with AOS versus without AOS or aphasia (Basilakos et al., 2017). More important, using clinically judged AOS presence versus absence as the diagnostic gold standard, the ASRS-1 has distinguished patients with PPAOS or PAOS with aphasia from patients with primary progressive aphasia (PPA) without AOS (Botha et al., 2015; Clark et al., 2015; Duffy et al., 2017; Josephs et al., 2012; Strand et al., 2014; Tetzloff et al., 2018). Moderate-to-strong correlations have been reported between the ASRS-1 and independent judgments of the presence and severity of PAOS, as well as acoustic measures that distinguish AOS from aphasia (Basilakos et al., 2017; Duffy et al., 2017; Josephs et al., 2013; Strand et al., 2014). Its capacity to detect worsening of AOS has been shown in patients with PAOS, concurrent with other clinical, acoustic, and neuroimaging measures of decline (Duffy et al., 2015; Josephs et al., 2014; Tetzloff et al., 2018). Some studies of stroke-related AOS have used the ASRS as the standard for AOS diagnosis (Basilakos et al., 2015, 2017; Moser et al., 2016).

The results of an unpublished study (Clark et al., 2016) that analyzed ASRS-1 results in people with isolated or varying combinations of neurodegenerative AOS, aphasia, and dysarthria led to a revision (ASRS-2) that eliminated three items and revised item description or scoring criteria for several items. Subsequent assessment of 20 patients yielded a strong correlation between the two ASRS versions, nonsignificant differences between their Total scores, excellent discrimination by Total ASRS-2 scores between patients with versus without a clinical diagnosis of AOS and a strong correlation between the ASRS-2 and clinical ratings of AOS severity.

Subsequent revisions to reduce scoring ambiguities and improve reliability, as well as reordering of items to group phonetic versus prosodic features, led to another version, the ASRS-3, whose rating form and scoring guidelines underwent further minor refinements over time to improve wording clarity and scorer reliability, leading eventually to the current version, the ASRS-3.5. These refinements did not alter the rating scale or change item substance. Therefore, the data from all versions of the ASRS-3, although mostly from the ASRS-3.5, were included in this study. For simplicity, we will refer to the ASRS-3.5 for the remainder of this article. The use of this version established that items capturing phonetic versus prosodic features are consistent with gestalt clinical judgments about the predominance of phonetic versus prosodic features in PPAOS; imaging pattern correlates were also identified (Utianski, Duffy, Clark, Strand, Botha, et al., 2018). Other studies have confirmed that, like the ASRS-1, the ASRS-3.5 is sensitive to the presence of PAOS and its progression, and that it correlates with other indices of AOS severity (Clark et al., 2021; Seckin et al., 2020; Utianski, Duffy, Clark, Strand, Boland, et al., 2018; Utianski et al., 2021). Its validity for use with non-degenerative causes of AOS (mostly stroke) as a composite index of AOS severity that correlates with other indices of severity (e.g., sound production accuracy, intelligibility, and acoustic measures) is supported by findings from research groups not involved in the measure's development (Basilakos et al., 2017; Haley et al., 2019; Wambaugh et al., 2019). Readers interested in the changes that have been made to the ASRS over time can compare the rating forms for the ASRS-1 (Strand et al., 2014), the ASRS 3.0 (Utianski, Duffy, Clark, Strand, Botha, et al., 2018), and the ASRS-3.5 (see the Appendix in this article).

Reliability

Good intrarater and interrater reliability have been reported for the Total score and item ratings for the ASRS-1 for patients with PPAOS, PAOS, and PPA (Duffy et al., 2015; Strand et al., 2014) and for patients with stroke-related AOS and aphasia by a research group not involved in the scale's development (Basilakos et al., 2015, 2017; Moser et al., 2016). More pertinent to this study, high interrater reliability has been reported for ASRS-3 Total scores in patients with PPAOS and PAOS (Utianski, Duffy, Clark, Strand, Botha, et al., 2018; Utianski et al., 2020). In addition, excellent interrater reliability for ASRS-3 Total scores for participants with stroke-related AOS and aphasia has been reported (Wambaugh et al., 2019); in that study, intraclass correlation coefficients (ICCs) were significant for 12 of the 13 items, but item reliability was poor for four items, leading to a conclusion that clearer rating guidelines may be needed to improve interrater reliability for some items. Relatedly, a recent study of the ASRS-2 (Hybbinette et al., 2021) examined reliability

among five speech-language pathologists without AOS expertise and limited ASRS training, for ratings of 10 post-stroke patients with AOS and aphasia, some of whom also had dysarthria. Intrajudge reliability for the Total score was, on average, moderate; item agreement was variable but generally in the moderate or good range. Interjudge reliability for the Total score and items was variable but, on average, poor. The authors appropriately concluded that reliability was not satisfactory.

To summarize, the item content of the ASRS and the rating of AOS feature prominence support its face validity and content validity for quantifying the presence and severity of AOS. Several studies demonstrate that its Total score distinguishes speakers with AOS from neurotypical speakers and individuals with aphasia who do not have AOS, that it correlates with other perceptual and acoustic measures of AOS presence and severity, and that it is sensitive to worsening of speech in PAOS. These data support the scale's construct and concurrent validity. Total score and item-level reliability have been poor in one study, but across several other studies, reliability has been good for the Total score and generally good or acceptable for most, but not all, item scores.

Some validity and reliability issues remain. One relates to the scale's inclusion of several features felt to overlap between AOS and aphasia and dysarthria, a possible vulnerability that could threaten its internal validity. Although previous data suggest that the scale does separate AOS from aphasia without AOS, the possible overlap with dysarthria has not been adequately assessed. In addition, a thorough item analysis for both validity and reliability purposes has not been conducted. In addition, from a practical diagnostic perspective, Total score values for determining AOS presence versus absence have not been established for the current 13-item version. Finally, given concerns about reliability of ratings for some items and the Total score, the effect of recent modifications in rating guidelines on interrater reliability require assessment.

The primary purposes of this study were to examine various aspects of validity and reliability for the current version of the ASRS, the ASRS-3.5, using data from a large patient cohort with varying combinations of neurodegenerative AOS, aphasia, and dysarthria. Using detailed item and Total score analyses, we determined (a) interrater reliability for items and the ASRS Total score; (b) the relationships among the ASRS items and their relationship to the Total score; (c) the degree to which the ASRS distinguishes patients with PAOS or PPAOS from patients with PPA without AOS and patients with dysarthria without AOS; (d) the relationship between the ASRS and independent ratings of AOS severity, articulatory errors, and intelligibility; (e) the relationship between the ASRS and clinical ratings of aphasia and dysarthria severity; (f) the relationship between the ASRS and years post onset and

age; and (g) Total score values for optimally identifying the presence versus absence of AOS.

Method

Participants

The entire study cohort included 308 participants drawn from two groups whose members had provided written consent to participate in IRB-approved NIH-funded research studies between July 2010 and December 2021. Unless otherwise specified, the data from both groups were eligible for assessment of ASRS reliability and validity.

One group included 218 individuals who were seen during their initial visit for studies of people with suspected isolated or predominant neurodegenerative AOS and/or aphasia; one additional individual was excluded because consensus could not be reached about whether AOS was present. Dysarthria could be present but must have been less severe than any AOS or aphasia. A number of these participants contributed to studies using earlier versions of the ASRS that were reviewed in the introduction; for this study, their recordings were rescored using the ASRS-3.5. No member of this group met criteria for a definitive non–speech-language neurological diagnosis (e.g., progressive supranuclear palsy [PSP]) at the time of their initial assessment. This group will be referred to as the "PPA/PAOS Group."

The second group included 90 individuals who were seen during their initial visit for studies of people with probable or possible PSP; they will be referred to as the "PSP Group." Many of them had dysarthria (79%), but few had AOS (8%) or aphasia (6%), although comprehensive language examination or rating of aphasia presence and severity was obtained for only one individual. These participants were among those in a recent study of motor speech disorders associated with PSP; see Clark et al. (2021) for a complete description of their characteristics. The inclusion of this group enhanced the assessment of the degree to which the ASRS is primarily sensitive to AOS and not dysarthria. PSP was considered an appropriate condition for this purpose, because dysarthria is a supportive diagnostic feature for PSP (Höglinger et al., 2017) and because many patients with PAOS eventually develop a dysarthria type like that encountered in PSP; in fact, many evolve to a PSP-like syndrome or later meet diagnostic criteria for possible or probable PSP (Duffy et al., 2014; Josephs et al., 2014).

Demographic characteristics and the distribution of primary speech and language findings for the entire cohort and participants with and without AOS are summarized in Table 1. For the entire cohort, median age was 70 years and 52% were female. Median disease duration in those with and without AOS was 3 years. Thirty-nine percent of

Table 1. Demographic and clinical features of participants with and without apraxia of speech (AOS) and the total study cohort.

Variable[a]	AOS absent (n = 188)	AOS present (n = 120)	Total cohort (n = 308)
Age	69 (63, 74)	71 (63, 76)	70 (63, 75)
Sex			
F	96 (51%)	63 (52%)	159 (52%)
M	92 (49%)	57 (48%)	149 (48%)
Handedness (R)	157 (86%)	107 (91%)	264 (88%)
Race (White)[b]	180 (96%)	116 (97%)	296 (96%)
Ethnicity (not Hispanic or Latino)	187 (99%)	119 (99%)	306 (99%)
Years from onset	3 (2, 4)	3 (2, 5)	3 (2, 5)
ASRS Total score (0 = no abnormalities)	2 (0, 5)	16 (12, 23)	6 (2, 14)
AOS severity (1–4; 4 = severe)	NA	2 (1, 2)	NA
AOS predominant subtype			
Phonetic	NA	49 (41%)	NA
Prosodic	NA	44 (37%)	NA
Mixed	NA	27 (22%)	NA
Dysarthria present[c]	68 (36%)	37 (31%)	105 (34%)
Dysarthria severity (1–4; 4 = severe)	2 (1, 2)	1 (1, 2)	2 (1, 2)
MSDSR (1–10; 10 = no impairment)	10 (8, 10)	6 (5, 7)	8 (6, 10)
Articulatory Error Score (AES; % errors)	9 (4, 26)	25 (11, 48)	16 (7, 41)
Aphasia present	103 (55%)	59 (49%)	162 (53%)
Aphasia severity	2 (1, 2)	1 (1, 2)	2 (1, 2)
WAB aphasia quotient	86 (78, 93)	96 (87, 98)	92 (83, 96)
Neurologic diagnosis			
PPAOS (i.e., no aphasia)	NA	46 (38%)	46 (15%)
PPA, agrammatic + AOS	0 (0%)	58 (48%)	58 (19%)
PPA, agrammatic, no AOS	8 (4%)	NA	8 (3%)
PPA, logopenic	51 (27%)	0 (0%)	51 (17%)
PPA, semantic	23 (12%)	0 (0%)	23 (7%)
PPA - Other[d]	17 (9%)	1 (1%)	18 (6%)
Other[e]	10 (5%)	9 (8%)	19 (6%)
PSP	79 (42%)	6 (5%)	85 (28%)

Note. Data are *n* (%) or median (interquartile range, Q1, Q3). F = female; M = male; ASRS = Apraxia of Speech Rating Scale; NA = not applicable; MSDSR = Motor Speech Disorders Severity Rating; WAB = Western Aphasia Battery; PPAOS = primary progressive apraxia of speech; PPA = primary progressive aphasia; PSP = progressive supranuclear palsy.

[a]Data were missing (*n*) for the following variables: handedness (8); race (1); years from onset (8); MSDSR (3). In the PSP Group, the WAB AQ and AES were rarely obtained. For those in the PSP Group for whom there was no complaint, suspicion, or informal observational evidence of aphasia (most of the group), aphasia was imputed as not present for the purpose of data analysis. [b]Non-White participants included American Indian/Alaska native, Asian Indian, Middle Eastern, Asian Chinese, and Asian Middle Eastern. [c]Hypokinetic, spastic, and mixed hypokinetic–spastic dysarthria accounted for > 90% of dysarthria types in the total cohort. [d]PPA not clearly meeting criteria for a single primary PPA variant. [e]Includes conditions such as corticobasal syndrome, behavioral variant frontotemporal dementia, posterior cerebral atrophy, dementia with Lewy bodies, possible PPAOS, AOS + other neurological symptoms, functional neurological disorder, no neurological disorder, and diagnosis not yet determined.

participants had AOS; median clinically rated AOS severity was moderate. Among the participants with AOS, 38% were classified as PPAOS (i.e., no evidence of aphasia). Dysarthria was present in 31% of participants with AOS (median dysarthria severity = mild) and 36% of patients without AOS (median dysarthria severity = moderate). Aphasia was present in 53% of the entire cohort, including in 49% of those with AOS and 55% of those without AOS; median aphasia severity was mild in those with AOS and moderate in those without AOS.

Test Measures

ASRS

The Appendix contains the updated ASRS-3.5 rating form, which includes its 13 items (features), guidance for rating items on a 5-point scale (in which 0 represents the absence of a feature and 1–4 represent degrees of prominence or severity of a feature when present), and additional guidance for rating some task-specific items. Tasks for eliciting a sufficient speech sample are provided at the end of the form. Items are grouped according to phonetic features (Items 1–4), prosodic features (Items 5–8), and other features that are either task specific or are not otherwise identified as phonetic versus prosodic in character (Items 9–13). The sum of the item scores within each of these groupings represent component scores (phonetic, prosodic, and other), which can be examined for their prominence. All item features have been noted in numerous studies of PAOS or AOS in general. The sum of item ratings yields a Total score, which can range from 0 to 52, where 0 = *no abnormal speech characteristics*.

Some of the ASRS features have been considered unique to AOS, but others—known as overlap features (McNeil et al., 2009)—may also be evident in aphasia or dysarthria. Table 2 identifies the items in which this overlap may exist on an a priori basis based on the prevailing AOS literature. The degree to which the data from this study are consistent with a priori designations of features as overlapping or not overlapping will be addressed in the Discussion section.

The ASRS items do not capture all perceptually identifiable features of AOS. This reflects a desire to minimize the burden an exhaustive list would impose on those who desire a relatively simple metric to support an AOS diagnosis and estimate of its severity, while at the same time including enough features to capture the gestalt of the abnormal speech pattern. Because the scale was initially developed to aid studies of PAOS and PPA, item selection was biased toward features we had observed in patients with those disorders. In fact, one feature (Item 11, reduced words or syllables per breath group), while documented in PAOS (Duffy, 2006; Josephs et al. 2012), has not been highlighted as a feature of more widely studied stroke-related AOS (cf. Takakura et al., 2019).

Apraxia of Speech Presence and Severity Rating (AOS Severity)

Independent of ASRS ratings, AOS presence and severity were rated on a 5-point scale, in which 0 represents the *absence of AOS* and 1–4 designate *mild, moderate, marked, and severe AOS*. This rating served as the gold standard for classifying participants as having AOS, or not, and for examining the relationship between the ASRS Total score and clinically rated AOS severity. In addition, a clinical judgment was made, independent of ASRS ratings,

about whether AOS was predominantly phonetic versus prosodic in its pattern, or mixed. Good reliability for these judgments has been reported (Clark et al., 2021; Josephs et al., 2012; Strand et al., 2014; Utianski, Duffy, Clark, Strand, Botha, et al., 2018); consensus about the presence, severity, and type of AOS, independent of ASRS scores, was established for all PPA/PAOS Group participants.

Dysarthria Presence and Severity Rating (Dysarthria Severity)

Dysarthria presence and severity were rated using the same 5-point scale as described for the AOS Severity rating. This permitted examination of the degree to which the ASRS correlates with clinically rated dysarthria severity, with an expectation that any relationship would be less strong than that between the ASRS and AOS Severity. Dysarthria type was also noted. Good reliability for dysarthria severity judgments has been reported (Clark et al., 2021; Josephs et al., 2012), and for all PPA/PAOS participants in this study, consensus about the presence, severity, and type of dysarthria was established; good reliability for a segment of PSP Group participants has been reported (Clark et al., 2021).

Aphasia Presence and Severity Rating (Aphasia Severity)

Aphasia presence and severity were rated using the same 5-point scale as described for the AOS Severity rating. The judgment was based on results of a detailed language assessment, described elsewhere (Josephs et al., 2012), that established aphasia presence and severity, and the specific variant of PPA. The Aphasia Severity rating was moderately correlated ($r = -.67$) with the Western

Table 2. Summary of ASRS interjudge reliability and frequency of abnormal features.

Item	Feature	Interjudge reliability[a]	Frequency of feature abnormality (rating ≥ 1)[b]		
			AOS only (n = 42)	Aphasia only (n = 100)	Dysarthria only (n = 65)
1	Distortions	Good–excellent	**88%**	18%	**78%**[c]
2	Distorted substitutions	Moderate–good	**76%**	28%	3%
3	Distorted additions	Moderate–excellent	*45%*	3%	0%
4	Increased errors with length/complexity	Good–excellent	**90%**	27%	14%
5	Segmentation within words	Good–excellent	**81%**	4%	5%[c]
6	Segmentation across words	Good–excellent	**86%**	2%	8%[c]
7	Slow rate	Good–excellent	**90%**	2%	42%[c]
8	Lengthened segments	Poor–good	**67%**	8%	5%[c]
9	AMRs	Moderate–excellent	**52%**	12%	**52%**[c]
10	SMRs	Excellent	**93%**	27%[c]	26%[c]
11	Reduced words or AMRs per breath	Good–excellent	*34%*	0%	2%
12	Visible/silent groping	Moderate–excellent	*26%*	5%[c]	0%
13	Audible false starts/restarts	Good–excellent	**69%**	**51%**[c]	17%[c]

Note. See Appendix for complete item labels. ASRS = Apraxia of Speech Rating Scale; AOS = apraxia of speech; AMRs = alternating motion rates; SMRs = sequential motion rates.

[a]Descriptors as suggested by Koo and Li (2016), based on ICC 95% confidence intervals (see Figure 1). [b]Features occurring in > 50% of participants are in bold; those occurring in 20%–50% of participants are in italics. [c]Considered a priori as an overlap feature with AOS.

Aphasia Battery–Revised Aphasia Quotient (Kertesz, 2006; WAB-AQ; see Table 1) in the study participants and was selected, rather than the WAB-AQ, as the aphasia severity index so the correlations among AOS, dysarthria, and aphasia severity would be based on similar scales. This rating permitted examination of the degree to which the ASRS correlates with aphasia severity, with an expectation that any relationship would be less strong than that between the ASRS and AOS Severity. Good reliability for aphasia severity judgments has been reported (Josephs et al., 2012); agreement about the presence and severity of aphasia was established for all PPA/PAOS Group participants.

Motor Speech Disorders Severity Rating

A 10-point (1 = *nonvocal*; 10 = *normal speech*) Motor Speech Disorders Severity Rating (MSDSR; see Duffy, 2020, as modified from Hillel et al., 1989) was used to estimate the effect of any MSD (AOS or dysarthria) on clinician-judged intelligibility. Because the ASRS is not designed to explicitly rate intelligibility but is intended to serve as a measure of AOS severity, the MSDSR permitted assessment of ASRS predictive and concurrent validity. Good interjudge reliability for the MSDSR has been reported for patients with dysarthria and patients with PAOS with or without dysarthria (Clark et al., 2021; Josephs et al., 2012).

Articulatory Error Score

An Articulatory Error Score (AES) was derived as an index of sound-level errors that can reflect AOS or aphasic phonological errors and, less often, dysarthria. The score represents the percentage of 56 words in which articulatory errors (excluding distortions that do not cross phonemic boundaries) occur during repetition of words or sentences (the Appendix includes the stimuli; see Duffy et al. [2015] for scoring criteria); good interjudge reliability has been reported (Utianski, Duffy, Clark, Strand, Botha, et al., 2018). It was expected that the AES would correlate with the ASRS Total score, perhaps most strongly with phonetic feature items.

Nonverbal Oral Apraxia

To assess the relationship between the ASRS and nonverbal oral apraxia (NVOA), an eight-item, 32-point measure of NVOA was administered (Botha et al., 2014). A significant but not more than moderate correlation was anticipated given the well-recognized co-occurrence but not uncommon dissociation between NVOA and AOS (Botha et al., 2014).

ASRS Reliability

Interjudge reliability was examined for 27 participants whose basic demographics (44% female; M_{age} =

71 years; median disease duration = 4 years) were representative of the entire cohort. Eleven were selected from first research visit data, whereas for 16 participants, a follow-up visit (ranging from the second to seventh visit) was selected to capture a greater range of severity due to disease progression. Aphasia was present in 30%; median severity = mild. Dysarthria was present in 63%; median severity = mild. AOS was present in 48%; patients with a predominance or relatively equal mix of phonetic and prosodic abnormalities were represented. The median ASRS Total score was 3 (interquartile range = 2, 6) for those without AOS and 25 (interquartile range = 21, 29) for those with AOS; the latter was higher than the entire cohort with AOS, which was desirable, because it captured a broader range of scores. Median AOS severity was marked.

The three judges who independently scored the ASRS for the reliability cohort (H.C., R.U., J.D.) were represented in 27, 16, and 11 of the 27 reliability comparisons that were made. All judges had considerable experience with neurodegenerative speech-language disorders and the ASRS.

Data Analysis

Demographic and clinical measure data were described using medians and interquartile ranges (Q1, Q3) for continuous variables and counts and percentages for categorical variables. Plots of absolute scores and score differences between raters were used to visually examine ASRS interrater reliability. ICCs with 95% confidence intervals (CIs) were computed using a two-way random effects model to examine consistency between raters.

Bar plots for item ratings were created for the entire cohort and subgroups with varying combinations of AOS, dysarthria, and aphasia. Histograms were created to display the distribution of the number of ASRS features scored as present and the distribution of ASRS Total scores in participants with versus without a clinical diagnosis of AOS.

Pairwise Kendall rank correlations were used to examine inter-item and item-Total correlations with α = .05; item-Total correlations were corrected for spurious overlap. Pairwise Spearman rank correlations were used to examine relationships among the ASRS Total and Component scores (corrected for spurious overlap), other clinical measures of disorder presence and severity, age, and years post onset, with α = .05; ASRS Total and Component score correlations were corrected for spurious overlap.

Because intercorrelations among the items showed considerable collinearity—as expected because all items are presumably related to the same underlying construct (AOS)—a penalized logistic regression model was used to examine the relationship between item scores and prediction of AOS presence. Ridge regression was employed to retain all items in the analysis but shrink the effects'

contribution to AOS prediction, thus "sharing" or "spreading" effects across collinear predictors. Aphasia and dysarthria were included as variables in the model but not penalized because we wished to examine the full influence of their effects. Only PPA/PAOS Group data were used for the regression analysis because language was not formally assessed for most of the PSP Group. We then fit an unpenalized logistic regression model predicting AOS presence by ASRS Total score only, and from the predicted values from that model, we derived the probability of AOS being present for all possible ASRS Total scores. We also plotted sensitivity and specificity curves to illustrate the derivation of an optimal cutpoint to maximize sensitivity and specificity of the Total score and visualize the uncertainty associated with borderline Total scores. The logistic model includes an intercept term, which slightly weights accuracy toward correctly identifying negative cases because 61% of the sample did not have AOS.

Statistical analyses were completed using the statistical software R (R Core Team, 2020) Version 4.0.3 in conjunction with the glmnet package Version 4.1–3 and the irr package Version 0.84 (Friedman et al., 2010; Gamer et al., 2019).

Results

Reliability

Plots illustrating reliability comparisons for the 13 ASRS items are provided in Figures 1 and 2. Because each item is rated on a 5-point scale that captures absence versus presence (0–1) and prominence/severity when present (1–4), less than exact agreement on an item can range from 1 to 4 points. Figure 1 shows a clustering of scores with exact agreement for all items. Only 3% of ratings across 351 comparisons (27 patients × 13 items) differed

Figure 1. Plots of interrater reliability scores and intraclass correlations (with 95% confidence intervals) for each Apraxia of Speech Rating Scale (ASRS) item (rated on 0–4 scale) and the Total score. Rater Pairs 1 and 2 refer to the two rater pairings that occurred among the three judges. See Appendix for complete item labels. AMRs = alternating motion rates; SMRs = sequential motion rates.
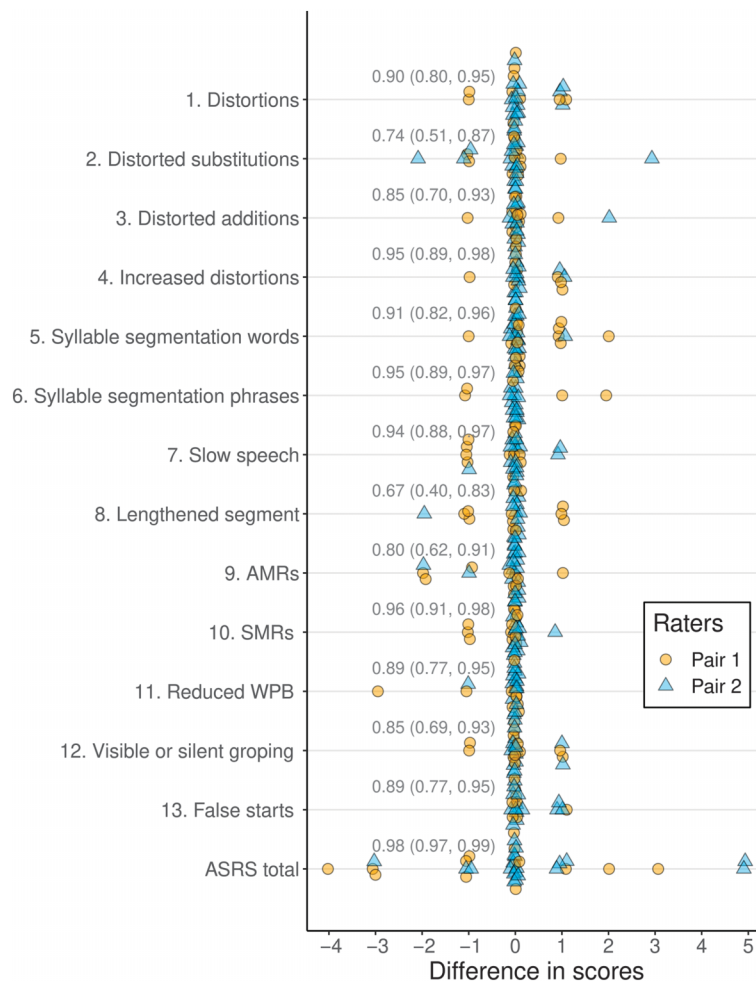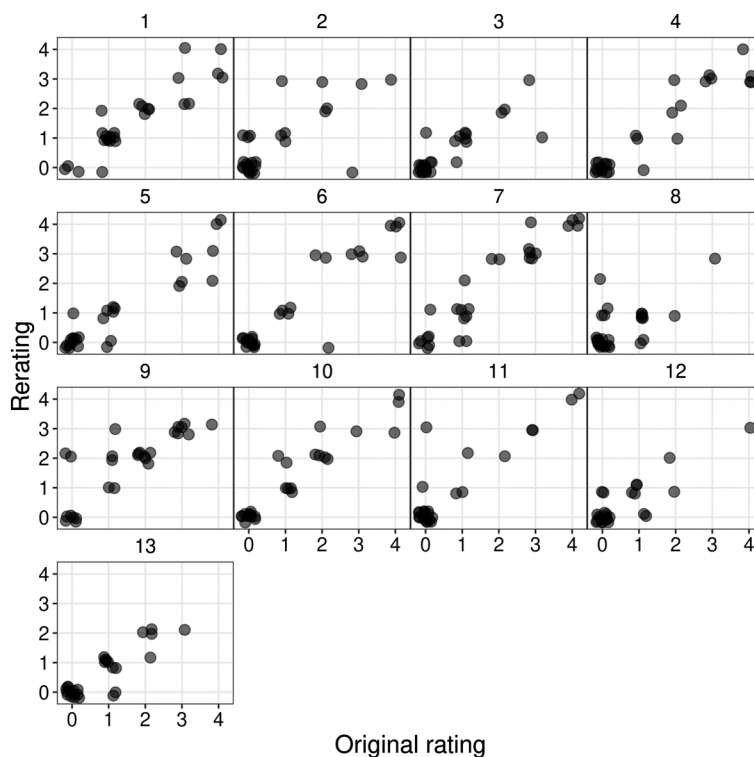
Figure 2. Plots of exact ratings between judges for each Apraxia of Speech Rating Scale (ASRS) item.

by more than 1 point. Figure 2, illustrating raw ratings between judges for each item, shows a largely linear relationship between raters, which suggests broadly adequate reliability across the range of abnormality that was evident for each item. The ICC estimates across the 13 items exceeded .90 for five items, ranged between .74 and .90 for seven items, and was .67 for one item (Item 8). Based on 95% CIs for the ICC estimates, reliability across the Items ranged from poor-good to excellent (see Table 2). Only for Item 8 did the lower boundary of its CI (0.40, 0.83) fall in the range of poor reliability (i.e., < .50); this will be addressed in the Discussion section. Thus, inter-judge item reliability fell in the moderate to excellent range (as per Koo & Li, 2016) for 12 of the 13 items.

Regarding the Total score, 37% of the Total score comparisons showed exact agreement and 70% were within 1 point; the maximum discrepancy was 5 points, which occurred for two (7%) of the 27 comparisons (see Figure 1). The original and rescored median Total ASRS scores were 7 and 8, respectively, each with an IQR of 3–24. The Total score ICC was excellent (.98; CI = .97, .99), higher than for any individual item.

Item and Total Score Distributions

Table 3 summarizes data for ASRS item ratings and the Total score for those with and without AOS, including data for a subgroup without AOS but with dysarthria or

aphasia and a subgroup with AOS but without dysarthria or aphasia. Variability is evident in the distribution of 0–4 ratings among items for those with and without AOS. Figure 3 shows bar plots that illustrate those differences for the entire cohort and its primary subgroups. The Figure 4 histogram shows the number of items scored as present (i.e., item rating ≥ 1) for participants without and with a clinical diagnosis of AOS.

AOS Absent

Within the entire group without AOS, for all 13 items, the most prevalent rating was 0 (feature not present), with frequency of 0 ratings ranging from 58% (Item 1) to 99% (Item 11). When a feature was present, for all items, it was most frequently rated 1 (infrequently present). The frequency of ratings > 1 was ≥ 10% for only three items (Items 1, 18%; 9, 10%; and 10, 12%).

The data for those with dysarthria or aphasia help identify the probable predominant source of abnormal scores. In those with dysarthria but no aphasia, frequency of abnormal ratings was considerably higher (> 20% higher) than for the entire group for Items 1 (distortions, 78% vs. 42%), 7 (slow rate, 42% vs. 18%), and 9 (alternating motion rates [AMRs], 52% vs. 29%). In those with aphasia but no dysarthria, abnormal ratings were somewhat more common (> 10%) than for the entire group for Items 2 (distorted substitutions, 28% vs. 17%) and 13 (audible false starts, 51% vs. 36%).

Table 3. Distribution of 0–4 ratings[a] (%) for ASRS items (see the Appendix for complete item labels); highest frequency for each item/group in boldface.

		AOS absent[b]			AOS present[c]	
Item	Rating	Total group (n = 188)	Dysarthria present aphasia absent (n = 65)	Aphasia present dysarthria absent (n = 100)	Total group (n = 120)	Dysarthria and aphasia absent (n = 42)
1. Distortions	0	**58**	22	**82**	8	12
	1	24	31	17	**34**	**43**
	2	12	**32**	1	32	29
	3	4	9	—	15	12
	4	2	6	—	11	5
2. Distorted substitutions	0	**83**	**97**	**72**	14	24
	1	16	3	27	**39**	**43**
	2	1	—	1	25	21
	3	—	—	—	11	2
	4	—	—	—	11	10
3. Distorted additions	0	**97**	**100**	**97**	**50**	**55**
	1	3	—	3	35	36
	2	—	—	—	10	5
	3	—	—	—	2	2
	4	—	—	—	2	2
4. Increased errors with length/complexity	0	**79**	**86**	**73**	4	10
	1	18	11	23	33	**40**
	2	3	3	4	**33**	29
	3	—	—	—	18	14
	4	—	—	—	12	7
5. Segmentation within words	0	**96**	**95**	**96**	24	19
	1	4	5	4	**34**	**43**
	2	—	—	—	26	26
	3	—	—	—	12	7
	4	—	—	—	4	5
6. Segmentation across words	0	**96**	**92**	**98**	18	14
	1	3	5	2	21	24
	2	1	3	—	21	17
	3	—	—	—	**30**	**33**
	4	—	—	—	11	12
7. Slow rate	0	**81**	**58**	**98**	11	10
	1	11	25	1	18	17
	2	6	14	1	17	19
	3	1	3	—	**32**	**36**
	4	—	—	—	22	19
8. Lengthened segments	0	**93**	**95**	**92**	42	33
	1	7	5	8	**49**	**57**
	2	—	—	—	8	10
	3	—	—	—	—	—
	4	—	—	—	—	—
9. AMRs	0	**72**	**48**	**88**	**37**	**48**
	1	18	29	12	25	21
	2	8	18	—	23	21
	3	2	3	—	12	10
	4	1	2	—	2	—
10. SMRs	0	**74**	**74**	**73**	8	7
	1	14	14	13	21	31
	2	10	12	10	**42**	**43**
	3	2	—	4	22	12
	4	—	—	—	8	7
11. Reduced words or AMRs per breath	0	**99**	**98**	**100**	**59**	**67**
	1	—	—	—	15	17
	2	1	2	—	14	10
	3	—	—	—	8	7
	4	—	—	—	4	—
12. Visible/silent groping	0	**96**	**100**	**95**	57	**74**
	1	4	—	5	33	24
	2	—	—	—	5	2
	3	—	—	—	4	—
	4	—	—	—	1	—

(table continues)

Table 3. *(Continued).*

| Item | Rating | AOS absent[b] | | | AOS present[c] | |
		Total group (*n* = 188)	Dysarthria present aphasia absent (*n* = 65)	Aphasia present dysarthria absent (*n* = 100)	Total group (*n* = 120)	Dysarthria and aphasia absent (*n* = 42)
13. Audible false starts/restarts	0	**64**	**83**	**49**	28	31
	1	29	12	43	**48**	**60**
	2	5	3	7	19	7
	3	2	2	1	3	2
	4	—	—	—	2	—
Total score *Mdn* (Q1, Q3)	0–52	2 (0, 5)	4 (2, 6)	1 (0, 4)	16 (12, 23)	14 (11, 22)

Note. Last row provides Total score median (Q1, Q3). Em dashes indicate no occurrence. ASRS = Apraxia of Speech Rating Scale; AOS = apraxia of speech; AMRs = alternating motion rates; SMRs = sequential motion rates.

[a]0 = *feature not present*; 4 = *always evident or marked in severity*. [b]In addition to the Dysarthria Present-Aphasia Absent and Aphasia Present-Dysarthria Absent subgroups shown here, the Total AOS Absent Group also contained participants with both aphasia and dysarthria (*n* = 3) and no aphasia or dysarthria (*n* = 20). Data for all subgroups are provided in Supplemental Material S1. [c]In addition to the dysarthria and aphasia absent subgroup shown here, the Total AOS Present Group also contained participants with both aphasia and dysarthria (*n* = 18), with aphasia but no dysarthria (*n* = 41), and with dysarthria but no aphasia (*n* = 19). Data for all subgroups are provided in the Supplemental Material S1.

Regarding the Total score (see Figure 5 and Table 1), although it ranged from 0 to 11, the median Total score of 2 (Q1, Q3 = 0, 5) is consistent with few items being rated as present and, when present, usually receiving a rating of 1. In fact, the median number of the 13 items rated as present was 2 (Q1, Q3 = 0, 3; see Table 4). Thus, Total scores were quite low for most participants with or without aphasia or dysarthria who did not have AOS.

Among the 188 participants without AOS, 20 (11%) had no evidence of aphasia or dysarthria. Their median ASRS Total score was 1 (Q1, Q3 = 0, 3). On only five items did more than two participants receive an abnormal rating, and only five items yielded any rating higher than 1 (data provided in Supplemental Material S1).

AOS Present

For the entire group with AOS, for only four items (3, 9, 11, and 12) did a majority or plurality of participants receive ratings of 0. The frequency of abnormal ratings across items ranged from 41% (Item 11) to 96% (Item 4). Ratings of 0 through 4 were given for all but Item 8, which had no ratings of 3 or 4.

In the Dysarthria and Aphasia Absent subgroup (in which all members had PPAOS), there were, in general, no substantial differences from the entire AOS Present Group in the distribution of 0–4 ratings across items. Ratings of 0 were somewhat more frequent (> 10%) in the Dysarthria and Aphasia Absent subgroup than the entire group for Items 9 (AMRs, 48% vs. 37%) and 12 (visible/silent groping, 74% vs. 57%).

The number of ASRS features scored as present between those with versus without AOS is informative (see Table 4). For those without AOS, the median number of features rated as present was 2, in contrast to those with mild AOS who had a median number of eight abnormal features; the median rose to 12 for those with marked and 13 for those with severe AOS. The histogram (see Figure 4) shows good separation between those with versus without AOS based on the number of items rated as abnormal. Thus, the number of abnormal features, independent of the ASRS Total score, was predictive of a clinical diagnosis of AOS.

The median Total score (see Table 1) for those with AOS was 16 (range: 5–43), considerably higher than the median of 2 for those without AOS (range: 0–11). The distribution of Total scores for those with versus without AOS (see Figure 5) shows good separation between the narrow range of low scores in those without AOS and the much broader distribution for those with AOS; the modal score for those with AOS falls at about the upper limit of the distribution for those without AOS. Overlap between the two groups occurred between Total scores of 5 and 11.

Inter-Item and Item-Total Score Correlations

All correlations among the 13 items (see Table 5) were positive and statistically significant; they ranged from weak (e.g., .10, between Items 9 and 13) to strong (e.g., .84, between Items 2 and 4), but were generally moderate (*Mdn* = .46). All item-Total score correlations were positive, significant, and, generally, moderately high (*Mdn* = .53; range: .34, Item 13, to .68, Item 4).

Item and Total Score Prediction of AOS Presence/Absence

A regression model that included all items as well as aphasia and dysarthria as predictors of a clinical diagnosis

Figure 3. Bar plot of ratings by item for participants: in the entire cohort; with versus without apraxia of speech (AOS); with AOS but no dysarthria or aphasia; with aphasia but no AOS or dysarthria; with dysarthria and no AOS or aphasia. PPAOS = primary progressive apraxia of speech.
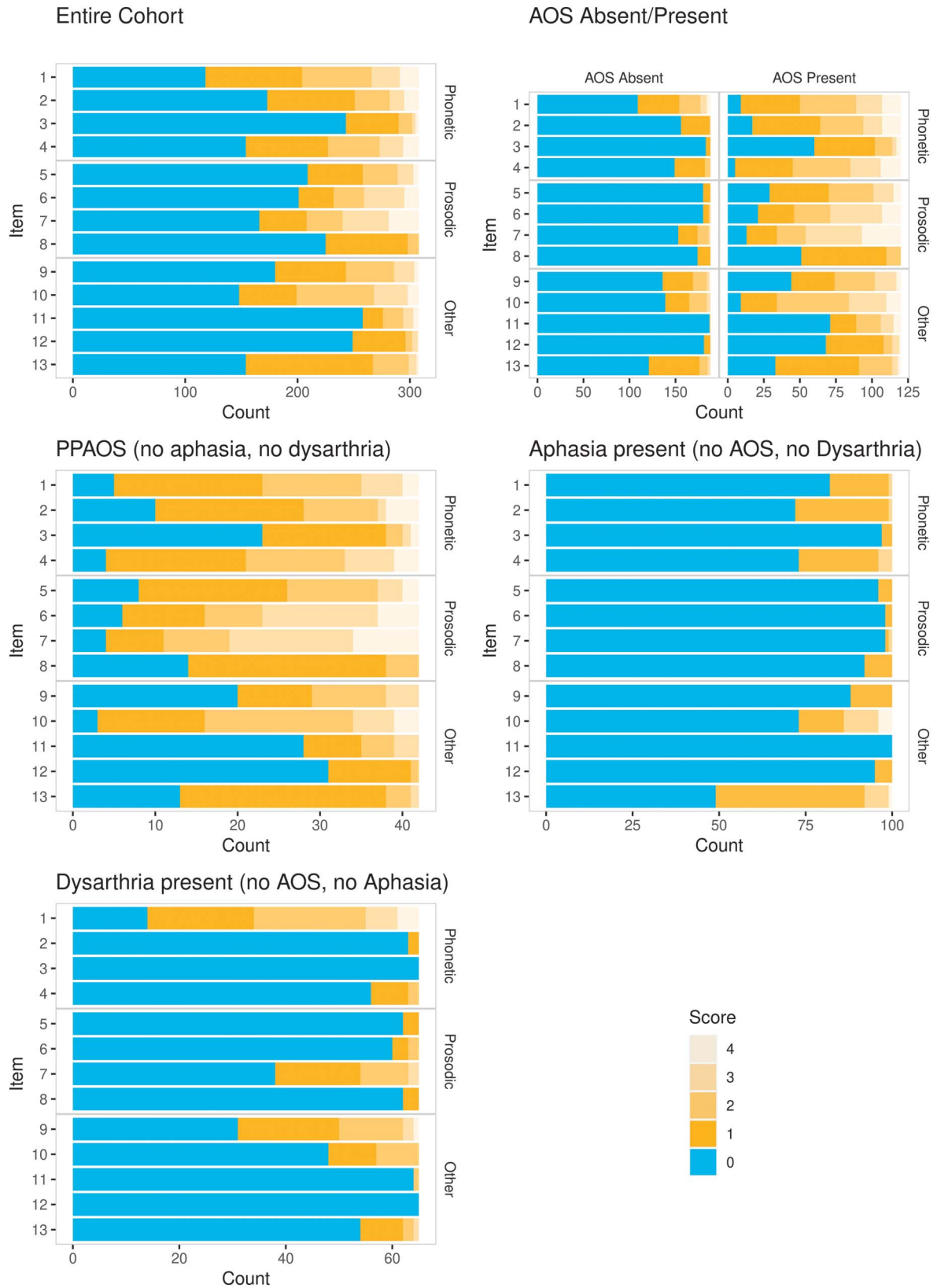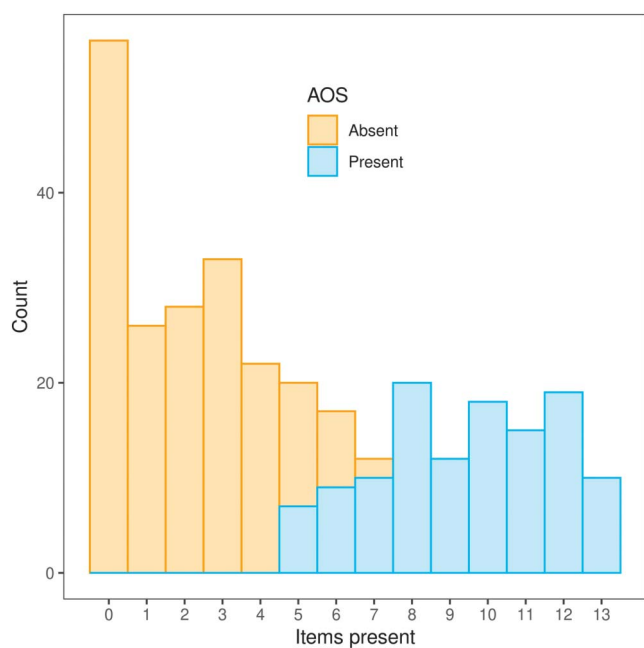
Figure 4. Distribution of number of items scored as present (rating ≥ 1) for participants without and with a clinical diagnosis of apraxia of speech (AOS).

of AOS (see Table 6) yielded item odds ratios ranging from 0.97 (Item 9) to 2.14 (Item 12); except for Item 9, all item odds ratios were 1.25 or higher. Thus, the presence of each of most of the abnormal features increased the

Figure 5. Distribution of Apraxia of Speech Rating Scale (ASRS) Total scores for participants with versus without apraxia of speech (AOS).
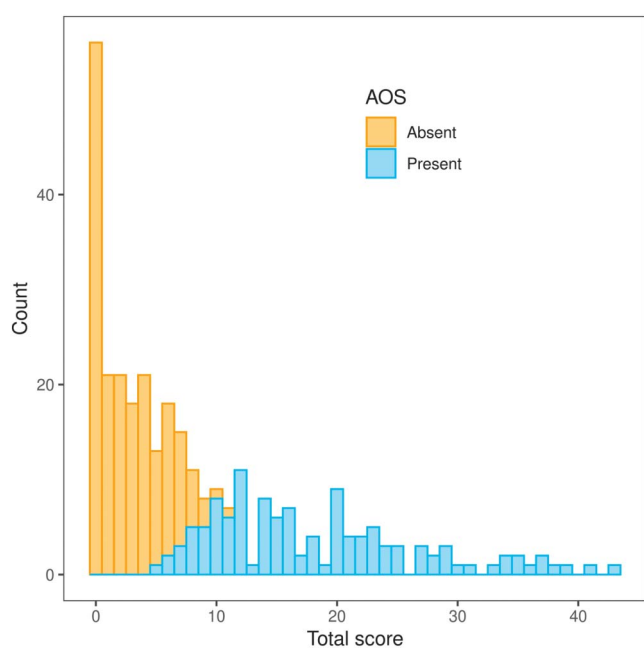
odds of AOS presence. In contrast, the < 1.0 odds ratios for aphasia (0.68) and dysarthria (0.50) indicate that the presence of either disorder was associated with a lower likelihood of AOS. Overall, the model resulted in an AUROC of 0.998 (95% CI [0.984, > .999]), nearly perfect separation of those with versus without a clinical diagnosis of AOS.

Figure 6 illustrates the sensitivity and specificity of the Total score using different values of the Total score as cutoffs. Also shown are the predicted probabilities of AOS presence from an unpenalized logistic regression model. The sensitivity and specificity curves intersect at a Total score of 8, the Youden's index that identifies the cutoff point that maximizes the quantity sensitivity plus specificity, treating false negatives and false positives as equally undesirable. From the logistic model, a cutoff of 9 would be best in terms of overall accuracy.

More simply put, our sample has slightly more AOS absent cases, and thus, overall accuracy (which is biased toward the sample makeup) is maximized at a cut point of 9, whereas sensitivity plus specificity (treating false positives and negatives equally) results in a cut point of 8; we also note that there is only one single more misclassification in our sample when using a cut point of 8 versus 9. We can summarize these results by saying the AOS probability curve shows that Total scores of 6 or lower indicate a > 90% likelihood that AOS is absent and Total scores of 11 or higher indicate a > 90% likelihood that AOS is present. In contrast, Total scores of 7–10, which were relatively uncommon in the data set (14%), are associated with probabilities for AOS presence that ascend from 22% to 85%, a range that would not support high diagnostic confidence.

ASRS Total and Component Score Relationships With Other Variables

Table 7 summarizes the Spearman correlations among the ASRS Total score, its component scores, and other variables. The Total score was strongly correlated with the 0–4 clinical rating of AOS severity (.70), a relationship that is demonstrated in Table 4. The Total score was moderately correlated with the AES (.43), an index of sound-level error severity; it was strongly correlated with the MSDSR (−.60), an index of intelligibility. In contrast, the Total score was not significantly correlated with clinical ratings of aphasia severity (−.07) or dysarthria severity (.04); it was weakly correlated with years post onset (.08) and age (.09).

Regarding ASRS component scores, the prominence/severity of phonetic, prosodic, and other components were all moderately or strongly correlated with the Total score, with each other, and with the AOS severity rating and MSDSR. Each was also correlated with the AES, most strongly for the phonetic component (0.52) and comparatively weakly for the prosodic component (0.27). Most of

Table 4. ASRS total score and number of items rated as abnormal in relation to clinical ratings of AOS presence/severity, median, and inter-quartile range (Q1, Q3).

	Clinical AOS presence/severity rating				
ASRS	0 (no AOS) (*n* = 188)	1 (mild; *n* = 60)	2 (moderate; *n* = 30)	3 (marked; *n* = 22)	4 (severe; *n* = 8)
Total score					
Mdn	2	12	20	28	36
Q1, Q3	0, 5	9, 14	15, 23	24, 32	34, 37
No. of items rated ≥ 1					
Mdn	2	8	10	12	13
Q1, Q3	0, 3	6, 10	9, 11	11, 12	12, 13

Note. AOS = apraxia of speech; ASRS = Apraxia of Speech Rating Scale.

the correlations between dysarthria and aphasia severity ratings with component scores were not statistically significant, and all were weak. Most of the correlations between age and component scores and other clinical ratings were not significant statistically, and all were weak. Years post onset was weakly or not significantly correlated with each of the component scores.

Discussion

The results of this large cohort study support the reliability and validity of the ASRS-3.5 as an index of AOS presence and prominence/severity of several of its important features. This discussion will address the primary findings that support this conclusion, as well as general clinical diagnostic considerations, caveats regarding use of the measure, and study and ASRS limitations.

Reliability

Overall, interjudge reliability for the ASRS items and Total score was good. Item reliability, as measured

by ICCs, was good or excellent for 12 of the 13 items (see Figure 1), and most item rating discrepancies were 1 point on the 5-point rating scale. These findings extend those for earlier versions of the ASRS and are generally consistent with reliability reported by some other research groups that have used the measure for patients with stroke-related AOS (summarized in the introduction).

The reliability for Item 8 (lengthened segments, independent of overall slow speaking rate) was poor-good (ICC 95% CI [0.40, 0.83]), less reliable than all other items and in keeping with its poor reliability reported by others (Hybbinette et al., 2021; Wambaugh et al., 2019). Two explanations for this seem plausible. First, determining the presence of a lengthened vowel or consonant segment can challenge perceptual judgment, especially when overall speech rate is slow, as is characteristic of AOS and many dysarthria types (future research using acoustic temporal measurements could help clarify this as well as support the validity of perceptual ratings for several other items/features). Second, in this study, Item 8 had less range in ratings than any other item among reliability participants; reduced range can artifactually reduce ICCs (Koo & Li, 2016). This reduced variability was also evident for Item 8

Table 5. Pairwise Kendall rank correlations among the 13 ASRS items (*N* = 308).

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	Total[a]
1		.47	.42	.51	.42	.42	.47	.31	.52	.50	.34	.34	.20	.53
2			.56	.84	.54	.50	.45	.42	.31	.61	.40	.50	.48	.61
3				.55	.58	.56	.49	.42	.35	.49	.45	.46	.25	.51
4					.59	.55	.50	.49	.39	.64	.41	.48	.46	.68
5						.79	.65	.54	.41	.55	.56	.41	.24	.62
6							.74	.56	.43	.55	.61	.40	.22	.63
7								.47	.48	.52	.54	.38	.12	.59
8									.31	.42	.38	.33	.25	.48
9										.43	.36	.29	.10	.45
10											.47	.41	.36	.64
11												.37	.11	.46
12													.37	.41
13														.34

Note. All item–item correlations are significant (α = .05).
[a]Item–total correlations are corrected for spurious overlap; all are significant (*p* < .001).

Table 6. Regression model odds ratios for item prediction of AOS presence, including dysarthria and aphasia as predictors (*N* = 308, including 120 with AOS).

Variable	Odds ratio
(intercept)	0.04
Item	
1	1.31
2	1.69
3	1.25
4	1.87
5	1.60
6	1.62
7	1.63
8	1.78
9	0.97
10	1.41
11	1.27
12	2.14
13	1.38
Dysarthria present	0.50
Aphasia present	0.68
Area under the ROC curve	0.998
(95% CI)	(0.984, > 0.999)

Note. AOS = apraxia of speech; ROC = receiver operating characteristic; CI = confidence interval.

in the entire study cohort, in which 97% of participants without AOS and 91% of those with AOS had ratings of 0 or 1, and no participant had a rating of 3 or 4. These findings suggest that the reliability challenge of Item 8 may lie in identifying the presence versus absence of the feature, rather than its predominance/severity when present. Despite this, Item 8 revealed distribution differences between participants with versus without AOS (see Tables 2 and 3), its moderate correlation with the Total score was comparable to that for other items (see Table 5), and

Figure 6. Sensitivity and specificity using different values of the Apraxia of Speech Rating Scale (ASRS) total score as cutoffs, as well as predicted probabilities of apraxia of speech (AOS) from an unpenalized logistic regression model.
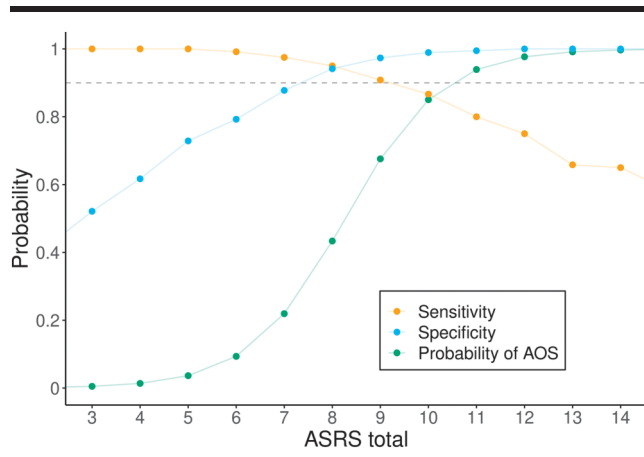
regression analysis revealed a positive odds ratio for its prediction of AOS presence (see Table 6). These latter findings, plus the item's descriptive value, argue against removing it from the ASRS for possible reliability weakness.

Reliability for the Total score was excellent (ICC = .98), superior to that for any individual item. This is reassuring for the reliability of an AOS diagnosis and estimate of severity based on the Total score. It suggests that Total score reliability is sufficiently robust for diagnostic and severity estimate purposes, even with lower reliability among some of its item scores.

These largely positive findings contrast with two studies that found less than satisfactory interjudge reliability for four items in one (Wambaugh et al., 2019) and, on average, poor reliability across all items in the other (Hybbinette et al., 2021). In the former study, raters experienced with AOS had minimal training with the ASRS, and in the latter study, ratings were completed by clinicians without AOS expertise and with limited ASRS training. Both studies suggested a need for clearer rating guidelines. Relative to those studies, the good reliability reported here may represent some combination of our considerable experience with neurodegenerative AOS and the ASRS, and the presumably improved rating guidelines that were added to the ASRS-3.5 for some items. In addition, it has not yet been established if reliability of ASRS ratings for patients with stroke-related AOS are intrinsically more challenging than for neurodegenerative AOS or, perhaps more likely, if other characteristics of our cohort, such as an absence of aphasia and dysarthria in a substantial number of cases, dysarthria never being worse than AOS in the PPA/PAOS Group, and AOS severity being only mild–moderate in most cases, made reliability easier for us to achieve; note, however, that reliable scoring for ASRS-1 and ASRS-3 Total scores has been reported for participants with stroke-related AOS (Basilakos et al., 2015, 2017; Moser et al., 2016; Wambaugh et al., 2019).

We are certain that experience in identifying the perceptual attributes of AOS and scoring the ASRS are important for its reliable scoring, if for no other reason than the long-recognized challenges that exist for perceptual judgments of AOS and motor speech disorders in general. Reliable scoring may take time and dedicated effort to achieve. Given the current and previous reliability findings, it is reasonable to conclude that the ASRS, at the least, *can be* reliably scored by clinicians and researchers experienced with people who have AOS whose aphasia or dysarthria, if present, are not more severe than AOS, and whose language and cognitive abilities are sufficient to provide an adequate sample of speech. To aid recognition of the abnormal features reflected in each ASRS item and facilitate reliability, the Supplemental Material

Table 7. Pairwise Spearman rank correlations among ASRS Total and Component scores[a] and other clinical variables (sample sizes vary because of unmeasured data).

Measure	ASRS phonetic	ASRS prosodic	ASRS other	AOS severity	AES	Dysarthria severity	MSDSR	Aphasia severity	Years post onset	Age
ASRS Total	.79	.72	.81	.70	.43	.04[b]	−.60	−.07[b]	.08	.09
ASRS phonetic		.55	.68	.60	.52	.14[b]	−.60	−.04[b]	.09	.03[b]
ASRS prosodic			.57	.47	.27	−.15[b]	−.56	−.23	.08[b]	.13
ASRS other				.61	.46	.08[b]	−.52	−.02[b]	.10	.08
AOS severity					.62	.22[b]	−.73	.02[b]	.30	.04[b]
AES						.18[b]	−.41	.13	.16	−.03[b]
Dysarthria severity							−.46	.27[b]	.13[b]	−.06[b]
MSDSR								.18	−.15	−.11
Aphasia severity									.04[b]	−.02[b]
Years postonset										.03[b]

Note. Correlations are significant at α = .05, unless otherwise indicated. ASRS = Apraxia of Speech Rating Scale; AOS = apraxia of speech; AES = Articulatory Error Score; MSDSR = Motor Speech Disorders Severity Rating.
[a]Correlations between ASRS Total and Phonetic, Prosodic, and Other component scores are corrected for spurious overlap. [b]Not significant ($p > .05$).

S2 contains a PowerPoint file with exemplars of the 13 features and some guidance for their scoring. We encourage those considering using the ASRS-3.5 to review that file when learning to recognize and rate the scale's abnormal features.

Finally, simply recognizing the presence of the abnormal features, without rating their prominence, is valuable diagnostically, given that the validity analysis showed that the number of abnormal features was a good predictor of AOS presence. Thus, identifying the number of abnormal ASRS features may be a compromise or alternative diagnostic marker if a clinician is not yet comfortable (reliable) rating feature prominence/severity.

Validity

Item Ratings (Feature Presence and Prominence in AOS, Dysarthria, and Aphasia)

All features represented by the 13 ASRS items were rated as present in participants with AOS, dysarthria, and aphasia, although with considerable differences in frequency and severity among the items and among the different subgroups, most importantly between those with versus without AOS (see summary in Table 2).

Ratings when AOS was absent. All 13 features were present in participants with aphasia and dysarthria, but, in general, their presence was neither common nor prominent, including for items identified a priori as overlapping between AOS and aphasia or dysarthria. That is, when AOS was absent, (a) the most prevalent rating for all items was 0 (feature not present); (b) when a feature was present, it was most often rated 1 (infrequent); and (c) the median number of the 13 items rated as present in those without AOS was only 2 (Q1, Q3 = 0, 3), although higher if dysarthria was present (*Mdn* = 4). It should be noted that some of the 20 participants in the AOS Absent group

who had neither aphasia nor dysarthria infrequently exhibited distortions, slow rate, mildly abnormal AMRs or SMRs (sequential motion rates), and occasional audible false starts (see Supplemental Material S1). Thus, although neurotypical speakers' Total scores should be quite low, they can be greater than 0.

Despite the usual low frequency of abnormal item ratings when AOS was absent, some ASRS items were sensitive to the presence of dysarthria or aphasia. In general, dysarthria appears to have driven abnormal ratings more than aphasia for items felt a priori to overlap between AOS and dysarthria, whereas aphasia tended to drive abnormal ratings for items felt a priori to overlap between AOS and aphasia. Exceptions to this and other issues related to overlap features between dysarthria and aphasia with AOS will be addressed in some detail under other observations. Overall, however, the data suggest that the sensitivity of ASRS items to dysarthria and aphasia is relatively limited. This is obviously a desirable trait for a measure primarily intended as an index of AOS presence and severity; it supports the construct and discriminative validity of the measure.

Ratings when AOS was present. Compared with the AOS Absent group, feature presence was notably more frequent and prominent for all 13 items in the AOS Present group, with frequency of abnormal ratings ranging from 41% (Item 11) to 96% (Item 4). Ratings of 0, 1, 2, 3, and 4 were given for 12 of the 13 items, which indicates that they vary in prominence across individuals with AOS, a desirable trait for a measure that purports to capture both the presence and prominence/severity of abnormal features.

There were no substantial differences in item score distributions between the entire AOS Present group and its Dysarthria and Aphasia Absent subgroup (AOS only) in which all members had PPAOS. The absence of a

feature was somewhat more prevalent ($\geq$ 10%) in the AOS only subgroup for Items 2 (distorted substitutions, 24% vs. 14%), 9 (AMRs, 48% vs. 37%), and 12 (visible/silent groping, 74% vs. 57%), likely reflecting some influence of aphasia (Item 2), or aphasia or dysarthria (Item 12), in the entire AOS Present group. The opposite pattern—a greater prevalence of 0 ratings in the entire AOS Present group—was not evident for any item.

Inter-Item and Item-Total Score Relationships

Correlations among the 13 items ranged from weak to strongly positive but were generally moderate (see Table 5). It is of interest, and additionally supportive of their validity, that the items with the greatest number of moderate or stronger correlations ($\geq$ 0.50) with other items are among those most often described as "core" features of AOS, distorted substitutions (Item 2), increased errors with length/complexity (Item 4), syllable segmentation within words and across words (Items 5 and 6), abnormal SMRs (Item 10). The overall moderate correlations are consistent with the expectation that all items would share some degree of variance because they all reflect speech or language deficits, particularly AOS, but not be so highly related that they are not capturing somewhat different aspects of the disorder(s).

Item and Total Score Prediction of AOS Presence

The odds ratios generated by the regression model that included each ASRS item plus aphasia and dysarthria as predictors of AOS reveal that the presence of each of 12 of the 13 features contributed to AOS diagnosis; Item 9 (AMRs) was essentially neutral in its contribution. In other words, the presence of each of the features to varying degrees increased the likelihood of an AOS diagnosis, regardless of whether dysarthria or aphasia was present. In contrast, the < 1.0 odds ratios for aphasia and dysarthria indicate that the presence of either disorder (dysarthria to a somewhat greater extent than aphasia) was associated with a lower likelihood of AOS presence. The combined results support a conclusion that the presence or absence of dysarthria and aphasia does not prevent the individual items' ability to predict AOS. This is confirmed by the AUROC of 0.998, indicative of nearly perfect separation of those with versus without a clinical diagnosis of AOS. These results support the predictive validity of the scale items.

Regarding Total score prediction of a clinical diagnosis of AOS, analysis established that a cutoff score of 8 maximized sensitivity (the proportion of those with clinically diagnosed AOS correctly identified by the ASRS as such) and specificity (the proportion of those with clinically diagnosed absence of AOS correctly identified as such); a logistic model that accounted for the greater number of participants without AOS in the sample, indicated that overall accuracy is maximized with a cutoff point of 9. The maximized sensitivity value of 0.95 and specificity value of 0.94 are excellent but, like any diagnostic measure based on behavior, not perfect. Rather than use a Total score of 8 (or 9) as a firm diagnostic cutpoint, it may be more appropriate to consider the probability of AOS being present within the range of scores in which overlap occurred between those with and without a clinical diagnosis of AOS in this study (see Figures 5 and 6). Although the data suggest that Total scores of 6 or lower are very unlikely to be associated with AOS ($\leq$ 10%) and that Total scores of 11 or higher are very likely ($\geq$ 90%) to be associated with AOS, Total scores of 7 to 10 fall in a diagnostic gray area in which the probability of AOS ranges from 22% (Total score = 7) to 85% (Total score = 10). Thus, scores in that range, more than scores outside of it, should encourage consideration of other information to inform diagnostic certainty, such as individual item ratings, the influence of other disorders (especially dysarthria or aphasia), acoustic measures, and independent clinical judgment. Similarly, but presumably less frequently, these other considerations may lead to conclusions that AOS is present when the Total score is less than 7 or that AOS is not present when the Total score is above 10. Finally, a gray area score may simply support a conclusion that a diagnosis of AOS is uncertain.

It is noteworthy that the number of features rated as present was also a good predictor of a clinical diagnosis of AOS. No individual without AOS received abnormal ratings on more than seven items. No individual with AOS had fewer than five items with abnormal ratings, and most had abnormal ratings on eight or more items (see Figure 4). Thus, as already noted, eight or more abnormal features may be an alternative diagnostic marker and 5, 6, or 7 abnormal items could be considered a gray area for AOS presence. This clinically useful finding supports the proposition that a diagnosis of AOS based on a single perceptual feature is not reasonable (Haley et al., 2019), but it requires modification of prior conclusions that the optimal subset of features necessary for an AOS diagnosis is unknown (Ballard et al., 2016) and that a checklist approach to AOS diagnosis may be difficult because of feature overlap with other disorders (McNeil et al., 2009). To the extent that the ASRS-3.5 items reflect a (less than exhaustive) cluster of features associated with AOS, it appears that the presence of eight or more of the 13 ASRS features is very supportive of an AOS diagnosis, at least in patients who fit within the characteristics of this study cohort.

Relatedly, we should note that this study did not investigate if fewer than 13 features/items can predict AOS presence without significant loss of sensitivity or specificity. Establishing the minimum number of features necessary for diagnosis—an ASRS short form—would reduce rating time but perhaps at the expense of descriptive value and reduced sensitivity to changes in severity

because the range of possible Total scores would be reduced. This is worthy of future investigation.

ASRS Total Score Relationship With Other Variables

The correlation between the ASRS Total score and the clinical rating of AOS severity (0.70) supports its convergent construct validity as a measure of AOS severity; Table 4 depicts this relationship in a way that can guide the conversion of Total scores to clinical labels of mild, moderate, marked, and severe levels of impairment. Convergent validity is also supported by correlations with two additional measures that capture aspects of AOS severity, the MSDSR (−0.60), an index of clinician-judged intelligibility, and the AES (0.43), a measure of percentage of sound-level errors during word and sentence repetition. In contrast, divergent construct validity (i.e., minimal relationship with measures of different constructs) is supported by nonmeaningful correlations with clinical ratings of aphasia severity (−0.07) and dysarthria severity (0.04), which suggest the ASRS is not sensitive to aphasia or dysarthria severity, at least at the generally mild to moderate levels of their severity in this study cohort. The Total score also was not meaningfully correlated with age (0.09) or years post onset (0.08), although the latter correlation is likely an artifact of the cross-sectional focus of this study in which ASRS data were drawn from the first research visit, which limited the range of AOS duration; longitudinal studies have shown that ASRS scores worsen over time (Utianski, Duffy, Clark, Strand, Boland, et al., 2018; Utianski et al., 2021).

Convergent and divergent construct validity is also supported by ASRS component score relationships. That is, the phonetic, prosodic, and other component subscores were each moderately or strongly correlated with the Total score, with each other, and with the AOS severity rating and MSDSR. The moderate correlation (.52) of the phonetic component with the AES (a measure of sound-level errors) and the comparatively weaker correlation of the prosodic component (.27) with the AES support the convergent validity of the phonetic subscore as an index of speech sound impairment and the relative divergence from that construct for the rate and prosodic features presumably tapped by the prosodic subscore. Like the Total score findings, correlations between component subscores and age and dysarthria and aphasia severity were weak. For reasons provided for the Total score, time post onset was also weakly correlated with each of the component scores.

Additional Observations

The item ratings permit a comparison of the frequency with which a set of abnormal speech features occur among AOS, dysarthria, and aphasia. Historically, there have been more frequent comparisons between AOS and aphasia than between AOS and dysarthria, and comparisons including all three disorders have been uncommon and with small participant numbers (Allison et al., 2020). If prevalence of a feature is the metric for comparison, and if 20% or higher prevalence—an admittedly arbitrary criterion—is considered as overlapping between AOS and dysarthria or aphasia, the data suggest that feature overlap is relative, with some features overlapping more than others and other features not overlapping at all; Table 2 identifies the degree of feature overlap for each of the 13 ASRS items.

Among the eight features considered a priori as overlapping between AOS and dysarthria, four met the 20% prevalence criteria for overlap in the dysarthria only subgroup: distortions (Item 1, 78%), slow rate (Item 7, 42%), AMRs, (Item 9, 52%), and SMRs (item 10, 26%). This is consistent with the well-recognized frequent presence of articulatory distortions, slow rate, and abnormal (typically slow) AMRs and SMRs in dysarthria, including the hypokinetic, spastic, and mixed hypokinetic–spastic types that predominated in this study. Although prevalence was higher for three of those four features in the AOS subgroup (prevalence was the same for AMRs), the data clearly support the validity of classifying distortions, slow rate, and abnormal AMRs and SMRs as overlapping between AOS and dysarthria. In addition, the frequent presence of abnormal AMRs in those with AOS and no dysarthria (52%) argues against a not-uncommon, oversimplified differential diagnosis teaching point that AMRs are normal in rate and phonetic accuracy in AOS and favors more nuanced guidance that AMRs are *less likely* to be abnormal than SMRs in AOS (e.g., 52% vs. 93% in this study). The remaining four features—within word segmentation (Item 5), between word segmentation (Item 6), lengthened segments (Item 8), and audible false starts/restarts (Item 13)—were rarely or uncommonly present in those with dysarthria alone but frequently were present in AOS alone (≥ 67%). At the least, this suggests that the "problem" of overlap with AOS for those features is insubstantial; that is, the probability is much higher that their presence reflects AOS than dysarthria, at least for dysarthric individuals meeting the entry criteria for this study. Finally, no feature considered a priori not to overlap between dysarthria and AOS met the 20% prevalence criteria for overlap, justifying their being characterized as nonoverlapping.

Among the three features considered a priori as overlapping between AOS and aphasia, two met criteria for overlap in the aphasia only subgroup: SMRs (Item 10, 27%), and audible false starts/restarts (Item 13, 51%). Although their prevalence was higher in AOS, the data do support the validity of classifying SMRs and audible false starts/restarts as overlapping between AOS and aphasia. The remaining feature, silent false starts/restarts (Item 12), was rarely present in aphasia (5%), suggesting that overlap

for that feature is insubstantial and that the probability is much higher that its presence reflects AOS than aphasia, at least for individuals with neurodegenerative, mild–moderate aphasia as in this study.

Among the 10 features considered a priori not to overlap between aphasia and AOS, eight failed to meet the 20% prevalence criteria for overlap, justifying their being characterized as nonoverlapping. However, although distorted substitutions (Item 2) and increases in them with increasing length/complexity (Item 4) have not traditionally been associated with aphasia (nondistorted substitutions and additions have been), studies have documented them in what are presumably aphasic phonological errors (e.g., Haley et al., 2019), including in patients with the logopenic variant of PPA (Haley et al., 2021). The current data support this; Items 2 and 4 features were both present in nearly 30% of participants with aphasia, although rarely were they rated as more than infrequently present; they were present in 76% and 90%, respectively, of those with AOS. The data thus suggest that Items 2 and 4 should be considered comparatively "weak" overlap features between AOS and aphasia.

In general, feature overlap between AOS and dysarthria, as reflected in their prevalence (see Table 2) and prominence (see Table 3), was somewhat more evident than overlap between AOS and aphasia. This provides some support for the proposition that the presumed speech planning/programming impairment reflected in neurodegenerative AOS lies closer to its network interface with motor speech control and execution than to its interface with language, particularly phonologic encoding. Although our cohort was diverse and had good representation across groups, we recognize that these observations about overlap features and their meaning might be altered for cohorts that differ in etiology or their distribution of AOS, dysarthria, and aphasia severity and subclassification features (e.g., relative prominence of AOS phonetic vs. prosodic features; type of dysarthria or aphasia).

Finally, comprehensive reviews have revealed that descriptions of features that justify an AOS diagnosis are often insufficient (Wambaugh et al., 2006), although they are improving (Ballard et al., 2015). The ASRS may provide an efficient way to describe speech characteristics in research participants with AOS. For example, studies could, at the least, indicate the number of ASRS features present (e.g., nine of 13 features) or, perhaps preferably, list the specific ASRS features that are present. This would provide more perceptual detail than vague descriptions of participants having "speech characteristics consistent with AOS."

Study Limitations and ASRS Limitations

The participants in this study had neurodegenerative diseases, so the results may not extrapolate to patients with AOS due to other conditions, such as stroke. It is noteworthy, for example, that 42 participants with AOS in this cohort had no evidence of aphasia or dysarthria; while this is a study strength, AOS without aphasia or dysarthria is uncommon when it results from stroke. In addition, AOS severity, and severity of aphasia and dysarthria when they were present, generally fell in the mild to moderate range, and when dysarthria was present with AOS, it was always less severe than AOS. These cohort characteristics could influence generalizability of reliability and validity findings, particularly for patients whose AOS and any co-occurring dysarthria and aphasia are severe. Indeed, as already noted, when AOS, aphasia or dysarthria are markedly severe many ASRS items may become unratable or a sufficient speech sample may be unobtainable. Thus, among patients we have followed longitudinally, few have received scores higher than the highest Total score of 43 in this study, and no patient has received the maximum possible Total score of 52. Rather, as severity increases, limited speech (including muteness), or patients' rejection of tasks as too difficult (e.g., SMRs, multisyllabic word repetition), can preclude valid rating of some or all items. Therefore, although some patients with a clinical rating of severe AOS can be validly rated (see Table 4), some cannot be, either because of AOS severity alone or because of the combined effects of AOS, dysarthria, aphasia, and, sometimes, other cognitive deficits. Nonetheless, the diagnostic and prognostic value of the ASRS for patients with progressive AOS may be greatest when AOS is mild or moderate, because that is the point at which the diagnosis and its implications for prognosis may be most important.

AOS is currently considered a disturbance of both articulation and prosody (McNeil et al., 2017). This study did not explore in any detail the degree to which the ASRS might permit a quantitative distinction between clinical judgments of phonetic versus prosodic predominant subtypes of neurodegenerative AOS, a distinction whose validity has received support from clinical disease course, neuroimaging, and underlying pathology levels of analysis (Josephs et al., 2021). Determining which ASRS phonetic and prosodic features best capture the clinical percept of phonetic versus prosodic predominance (when the distinction exists) could enhance ASRS validity and may establish a quantitative metric that supports the distinction. It would also aid investigations of imaging correlates and whether AOS subtypes are evident in patients with AOS from nondegenerative causes.

As already noted, our findings of good interrater reliability were based on ratings by clinicians experienced with AOS, aphasia, dysarthria, and the ASRS. When combined with findings of other studies, it is prudent to conclude that ratings can be reliable but may require dedicated effort to achieve. The guidelines in the ASRS-3.5

rating form and the examples provided in the Supplemental Material S2 should be helpful in this regard.

Although the cross-sectional data in this study showed only a weak correlation with years post onset at a first research visit, longitudinal data in other studies of neurodegenerative AOS have demonstrated that the ASRS is responsive to increasing AOS severity over time, generally at 1- to 2-year intervals (Duffy et al., 2015; Josephs et al., 2014; Tetzloff et al., 2018; Utianski, Duffy, Clark, Strand, Boland, et al., 2018; Utianski et al., 2021). Whether it is sensitive to improvement when some recovery is expected (e.g., stroke etiology) or in response to treatment is yet to be established.

The intended use of the ASRS is as a diagnostic tool and index of speech impairment. It does not directly assess intelligibility, although it is moderately correlated with intelligibility ratings. It was not designed to have social validity or measure communication participation, and its relationship with communication participation may be neither strong nor straightforward (Utianski et al., 2020).

These limitations and caveats support the continued importance of clinical judgment in the diagnosis of AOS. While the results of this study provide strong support for the ASRS as a clinical and research diagnostic aid, confidence in a diagnosis of AOS is enhanced when both clinical judgment and formal, structured measures yield converging results.

Data Availability Statement

Additional data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

References

Allison, K. M., Cordella, C., Iuzzini-Seigel, J., & Green, J. R. (2020). Differential diagnosis of apraxia of speech in children and adults: A scoping review. *Journal of Speech, Language, and Hearing Research, 63*(9), 2952–2994. https://doi.org/10.1044/2020_JSLHR-20-00061

Ballard, K. J., Azizi, L., Duffy, J. R., McNeil, M. R., Halaki, M., O'Dwyer, N., Layfield, C., Scholl, D. I., Vogel, A. P., & Robin, D. A. (2016). A predictive model for diagnosing stroke-related apraxia of speech. *Neuropsychologia, 81,* 129–139. https://doi.org/10.1016/j.neuropsychologia.2015.12.010

Ballard, K. J., Wambaugh, J. L., Duffy, J. R., Layfield, C., Maas, E., Mauszycki, S., & McNeil, M. R. (2015). Treatment for acquired apraxia of speech: A systematic review of intervention research between 2004 and 2012. *American Journal of Speech-Language Pathology, 24*(2), 316–337. https://doi.org/10.1044/2015_AJSLP-14-0118

Basilakos, A., Rorden, C., Bonilha, L., Moser, D., & Fridriksson, J. (2015). Patterns of poststroke brain damage that predict speech production errors in apraxia of speech and aphasia dissociate. *Stroke, 46*(6), 1561–1566. https://doi.org/10.1161/strokeaha.115.009211

Basilakos, A., Yourganov, G., den Ouden, D. B., Fogerty, D., Rorden, C., Feenaughty, L., & Fridriksson, J. (2017). A multivariate analytic approach to the differential diagnosis of apraxia of speech. *Journal of Speech, Language, and Hearing Research, 60*(12), 3378–3392. https://doi.org/10.1044/2017_JSLHR-S-16-0443

Botha, H., Duffy, J. R., Strand, E. A., Machulda, M. M., Whitwell, J. L., & Josephs, K. A. (2014). Nonverbal oral apraxia in primary progressive aphasia and apraxia of speech. *Neurology, 82*(19), 1729–1735. https://doi.org/10.1212/WNL.0000000000000412

Botha, H., Duffy, J. R., Whitwell, J. L., Strand, E. A., Machulda, M. M., Schwarz, C. G., Reid, R. I., Spychalla, A. J., Senjem, M. L., Jones, D. T., Lowe, V., Jack, C. R., & Josephs, K. A. (2015). Classification and clinicoradiologic features of primary progressive aphasia (PPA) and apraxia of speech. *Cortex, 69,* 220–236. https://doi.org/10.1016/j.cortex.2015.05.013

Clark, H. M., Duffy, J. R., Strand, E. A., & Josephs, K. A. (2016, March 5). *Revisions to the Apraxia of Speech Rating Scale* [Paper presentation]. Conference on Motor Speech.

Clark, H. M., Hanley, H., Duffy, J. R., Strand, E. A., & Josephs, K. A. (2015, May 30). *Speech features unique to apraxia of speech: preliminary item analysis of the Apraxia of Speech Rating Scale (ASRS)* [Paper presentation]. Clinical Aphasiology Conference.

Clark, H. M., Utianski, R. L., Ali, F., Botha, H., Whitwell, J. L., & Josephs, K. A. (2021). Motor speech disorders and communication limitations in progressive supranuclear palsy. *American Journal of Speech-Language Pathology, 30*(3S), 1361–1372. https://doi.org/10.1044/2020_AJSLP-20-00126

Duffy, J. R. (2006). Apraxia of speech in degenerative neurologic disease. *Aphasiology, 20*(6), 511–527. https://doi.org/10.1080/02687030600597358

Duffy, J. R. (2020). *Motor speech disorders: Substrates, differential diagnosis, and management* (4th ed.). Elsevier.

Duffy, J. R., Hanley, H., Utianski, R. L., Clark, H. C., Strand, E. A., Josephs, K. A., & Whitwell, J. L. (2017). Temporal acoustic measures distinguish primary progressive apraxia of speech from primary progressive aphasia. *Brain and Language, 168,* 84–94. https://doi.org/10.1016/j.bandl.2017.01.012

Duffy, J. R., Strand, E. A., Clark, H. M., 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*(2), 88–100. htpps://doi.org/10.1044/2015_AJSLP-14-0174

Duffy, J. R., Strand, E. A., & Josephs, K. A. (2014). Motor speech disorders associated with primary progressive aphasia. *Aphasiology, 28*(8–9), 1004–1017. https://doi.org/10.1080/02687038.2013.869307

Duffy, J. R., Utianski, R. L., & Josephs, K. A. (2021). Primary progressive apraxia of speech: From recognition to diagnosis and care. *Aphasiology, 35*(4), 560–591. https://doi.org/10.1080/02687038.2020.1787732

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22. https://www.jstatsoft.org/v33/i01/

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various coefficients of interrater reliability and agreement*. R package version 0.84.1. https://CRAN.R-project.org/package=irr

Haley, K. L., Jacks, A., de Riesthal, M., Abou-Khalil, R., & Roth, H. L. (2012). Toward a quantitative basis for assessment and diagnosis of apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55*(5), S1502–S1517. https://doi.org/10.1044/1092-4388(2012/11-0318)

Haley, K. L., Jacks, A., Jarrett, J., Ray, T., Cunningham, K. T., Gorno-Tempini, M. L., & Henry, M. L. (2021). Speech metrics and samples that differentiate between nonfluent/agrammatic and logopenic variants of primary progressive aphasia. *Journal of Speech, Language, and Hearing Research, 64*(3), 754–775. https://doi.org/10.1044/2020_JSLHR-20-00445

Haley, K. L., Smith, M., & Wambaugh, J. (2019). Sound distortion errors in aphasia with apraxia of speech. *American Journal of Speech-Language Pathology, 28*(1), 121–135. https://doi.org/10.1044/2018_AJSLP-17-0186

Hillel, A. D., Miller, R. M., Yorkston, K., McDonald, E., Norris, F. H., & Konikow, N. (1989). Amyotrophic lateral sclerosis severity scale. *Neuroepidemiology, 8*(3), 142–150. https://doi.org/10.1159/000110176

Höglinger, G. U., Respondek, G., Stamelou, M., Kurz, C., Josephs, K. A., Lang, A. E., Mollenhauer, B., Müller, U., Nilsson, C., Whitwell, J. L., Arzberger, T., Englund, E., Gelpi, E., Giese, A., Irwin, D. J., Meissner, W. G., Pantelyat, A., Rajput, A., van Swieten, J. C., ... Movement Disorder Society-endorsed PSP Study Group. (2017). Clinical diagnosis of progressive supranuclear palsy: The movement disorder society criteria. *Movement Disorders: Official journal of the Movement Disorder Society, 32*(6), 853–864. https://doi.org/10.1002/mds.26987

Hybbinette, H., Östberg, P., & Schalling, E. (2021). Intra- and interjudge reliability of the Apraxia of Speech Rating Scale in early stroke patients. *Journal of Communication Disorders, 89,* 106076. https://doi.org/10.1016/j.jcomdis.2020.106076

Josephs, K. A., Duffy, J. R., Clark, H. M., Utianski, R. L., Strand, E. A., Machulda, M. M., Botha, H., Martin, P. R., Trang Thu Pham, N., Stierwalt, J., Ali, F., Buciuc, M., Baker, M., Fernando De Castro, C. H., Spychalla, A. J., Schwarz, C. G., Reid, R. I., Senjem, M. L., Jack, C. R., Jr., ... Whitwell, J. L. (2021). A molecular pathology, neurobiology, biochemical, genetic and neuroimaging study of progressive apraxia of speech. *Nature Communications, 12*(1), 3452. https://doi.org/10.1038/s41467-021-23687-8

Josephs, K. A., Duffy, J. R., Strand, E. A., Machulda, M. M., Senjem, M. L., Gunter, J. L., Schwarz, C. G., Reid, R. I., Spychalla, A. J., Lowe, V. J., Jack, C. R., Jr., & Whitwell, J. (2014). The evolution of primary progressive apraxia of speech. *Brain, 137*(10), 2783–2795. https://doi.org/10.1093/brain/awu223

Josephs, K. A., Duffy, J. R., Strand, E. A., Machulda, M. M., Senjem, M. L., Lowe, V. J., Jack, C. R., Jr., & Whitwell, J. L. (2013). Syndromes dominated by apraxia of speech show distinct characteristics from agrammatic PPA. *Neurology, 81*(4), 337–345. https://doi.org/10.1212/WNL.0b013e31829c5ed5

Josephs, K. A., Duffy, J. R., Strand, E. A., Machulda, M. M., Senjem, M. L., Master, A. V., Lowe, V. J., Jack, C. R., Jr.,

& Whitwell, J. L. (2012). Characterizing a neurodegenerative syndrome: Primary progressive apraxia of speech. *Brain, 135*(5), 1522–1536. https://doi.org/10.1093/brain/aws032

Kertesz, A. (2006). *Western Aphasia Battery–Revised (WAB-R)*. APA PsycTests. https://doi.org/10.1037/t15168-000

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

McNeil, M. R., Ballard, K. J., Duffy, J. R., & Wambaugh, J. (2017). Apraxia of speech theory, assessment, differential diagnosis, and treatment: Past, present, and future. In P. van Lieshout, B. Maassen, & M. Terband (Eds.), *Speech motor control in normal and disordered speech: Future developments in theory and methodology* (pp. 195–221). American Speech-Language-Hearing Association.

McNeil, M. R., Robin, D., & Schmidt, R. (2009). Apraxia of speech. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (2nd ed., pp. 249–268). Thieme.

Moser, D., Basilakos, A., Fillmore, P., & Fridriksson, J. (2016). Brain damage associated with apraxia of speech: Evidence from case studies. *Neurocase, 22*(4), 346–356. https://doi.org/10.1080/13554794.2016.1172645

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Seckin, Z. I., Whitwell, J. L., Utianski, R. L., Botha, H., Ali, F., Duffy, J., Clark, H. M., Machulda, M., Jordan, L. G., Min, H.-K., Lowe, V. J., & Josephs, K. A. (2020). Ioflupane 123I (DAT scan) SPECT identifies dopamine receptor dysfunction early in the disease course in progressive apraxia of speech. *Journal of Neurology, 267*(9), 2603–2611. https://doi.org/10.1007/s00415-020-09883-4

Strand, E. A., Duffy, J. R., Clark, H. M., & Josephs, K. (2014). The Apraxia of Speech Rating Scale: A tool for diagnosis and description of apraxia of speech. *Journal of Communication Disorders, 51,* 43–50. https://doi.org/10.1016/j.jcomdis.2014.06.008

Takakura, Y., Otsuki, M., Sakai, S., Tajima, Y., Mito, Y., Ogata, A., Koshimizu, S., Yoshino, M., Uemori, G., Takakura, S., & Nakagawa, Y. (2019). Sub-classification of apraxia of speech in patients with cerebrovascular and neurodegenerative diseases. *Brain and Cognition, 130,* 1–10. https://doi.org/10.1016/j.bandc.2018.11.005

Tetzloff, K. A., Utianski, R. L., Duffy, J. R., Clark, H. M., Strand, E. A., Josephs, K. A., & Whitwell, J. L. (2018). Quantitative analysis of agrammatism in agrammatic primary progressive aphasia and dominant apraxia of speech. *Journal of Speech, Language, and Hearing Research, 61*(9), 2337–2346. https://doi.org/10.1044/2018_JSLHR-L-17-0474

Utianski, R. L., Clark, H. M., Duffy, J. R., Botha, H., Whitwell, J. L., & Josephs, K. A. (2020). Communication limitations in patients with progressive apraxia of speech and aphasia. *American Journal of Speech-Language Pathology, 29*(4), 1976–1986. https://doi.org/10.1044/2020_AJSLP-20-00012

Utianski, R. L., Duffy, J. R., Clark, H. M., Strand, E. A., Boland, S. M., Machulda, M. M., Whitwell, J. L., & Josephs, K. A. (2018). Clinical progression in four cases of primary progressive apraxia of speech. *American Journal of Speech-Language Pathology, 27*(4), 1303–1318. https://doi.org/10.1044/2018_AJSLP-17-0227

Utianski, R. L., Duffy, J. R., Clark, H. M., Strand, E. A., Botha, H., Schwarz, C. G., Machulda, M. M., Senjem, M. L., Spychalla, A. J., Jack, C. R., Jr., Petersen, R. C., Lowe,

V. J., Whitwell, J. L., & Josephs, K. A. (2018). Prosodic and phonetic subtypes of primary progressive apraxia of speech. *Brain and Language, 184,* 54–65. https://doi.org/10.1016/j.bandl.2018.06.004

Utianski, R. L., Martin, P. R., Hanley, H., Duffy, J. R., Botha, H., Clark, H. M., Whitwell, J. L., & Josephs, K. A. (2021). A longitudinal evaluation of speech rate in primary progressive apraxia of speech. *Journal of Speech, Language, and Hearing Research, 64*(2), 392–404. https://doi.org/10.1044/2020_JSLHR-20-00253

Wambaugh, J. L., Bailey, D. J., Mauszycki, S. C., & Bunker, L. D. (2019). Interrater reliability and concurrent validity for the Apraxia of Speech Rating Scale 3.0: Application with persons with acquired apraxia of speech and aphasia. *American Journal of Speech-Language Pathology, 28*(2S), 895–904. https://doi.org/10.1044/2018_AJSLP-MSC18-18-0099

Wambaugh, J. L., Duffy, J. R., McNeil, M. R., Robin, D. A., & Rogers, M. A. (2006). Treatment guidelines for acquired apraxia of speech: A synthesis and evaluation of the evidence. *Journal of Medical Speech-Language Pathology, 14*(2), xv–xxxiii.

Appendix (p. 1 of 2)

Apraxia of Speech Rating Scale 3.5

Patient Name:_____ #:_____ Date:_____ Examiner:_____

Score	0	1	2	3	4
Description	**Not observed in any task**	**Infrequent**	**Frequent but not pervasive**	**Very often evident but not marked in severity**	**Nearly always evident and/or marked in severity**
Guidelines (estimate of feature prevalence)	No more than one occurrence	Noted more than once (but less than about 20% of words)	Noted in about 20–50% of words	Noted in majority of words	Noted in nearly all words

Exceptions
- *Score no higher than "2" if present only during repetition tasks.*
- *Score a "4" if intelligibility is more than mildly reduced by the feature.*
- *Performance on AMRs and SMRs should be considered only for items 9–11*

Phonetic Features*

1 Sound distortions (excluding distorted substitutions or distorted additions)
2 Distorted sound substitutions
3 Distorted sound/syllable additions (including intrusive schwa)
4 Increased sound distortions or distorted sound substitutions with increased utterance length or increased syllable/word articulatory complexity

Prosodic Features*

5 Syllable segmentation within words > 1 syllable
(Brief silent interval between syllables &/or inappropriate equalized stress across syllables)
6 Syllable segmentation across words in phrases/sentences
(Increased inter-word intervals &/or inappropriate equalized stress across words)
7 Slow overall speech rate (apart from pauses for word retrieval and/or verbal formulation)
8 Lengthened vowel &/or consonant segments independent of overall slow speaking rate

Other*

9 **RATE ONLY FOR AMRs** (alternating motion rates, as in rapid repetition of "puh puh puh")**:** Slow &/or off-target (in place, manner, &/or voicing)
0 = AMRs normal; 1 = rare and mild, 2 = frequent but mild; 3 = moderate, 4 = severe
10 **RATE ONLY FOR SMRs** (sequential motion rates, as in rapid repetition of "puh tuh kuh"): Slow (gaps within sequences), segmented (gaps between sequences), incorrectly sequenced, &/or off-target (in place, manner, and/or voicing)
0 = SMRs normal; 1 = any one of the listed features, 2 = any two of the listed features;
3 = any three of the listed features, 4 = four of the listed features
11 One or both of the following: Reduced words per breath group during phrase/sentence production relative to maximum vowel duration; reduced # of AMR repetitions per breath group in the absence of decreased respiratory capacity.
Score on estimated average number of syllables/repetitions per breath group across tasks:
0 = more than 7; 1 = 6–7; 2 = 4–5; 3 = 3–4; 4 = 2 or less
12 Silent articulatory false starts/restarts or groping
13 Audible false starts/restarts or groping, including sound repetitions, excluding fillers and unambiguous semantic false starts (e.g., spoo…fork)

*Note that some of these features can also be present in individuals with aphasia &/or dysarthria. The degree to which feature overlap occurs can range from infrequent to frequent. Overlap frequency may vary as a function of severity, type of aphasia or dysarthria, or etiology (e.g., stroke versus neurodegenerative disease).

Total of Ratings:_____

Apraxia of Speech Rating Scale 3.5

Ratings should be based on speech samples that include:

- Brief conversation
- Solicited narrative (e.g., picture description) or oral reading of a short paragraph
- Supplemental tasks

    o Prolonged Vowel
    o Speech AMRs (If necessary, reinstruct and score the final effort)
    o Speech SMRs (If necessary, reinstruct and score the final effort)
    o Word Repetition (3 consecutive repetitions of each word)

        ▪ Cat
        ▪ Catnip
        ▪ Catapult
        ▪ Catastrophe
        ▪ Snowman
        ▪ Artillery
        ▪ Stethoscope
        ▪ Rhinoceros
        ▪ Volcano
        ▪ Harmonica
        ▪ Specific
        ▪ Statistics
        ▪ Aluminum

    o Sentence Repetition (1 trial)

        ▪ We saw several wild animals.
        ▪ My physician wrote out a prescription.
        ▪ The municipal judge sentenced the criminal.

Apraxia of Speech Rating Scale 3.5.2 2022

Note. Adapted with permission from Josephs et al. (2012) and Strand et al. (2014).