



A preliminary validation of a dynamic speech motor assessment for Swedish-speaking children with childhood apraxia of speech

Susanne Rex, Anders Sand, Edythe Strand, Kristina Hansson & Anita McAllister

To cite this article: Susanne Rex, Anders Sand, Edythe Strand, Kristina Hansson & Anita McAllister (2022) A preliminary validation of a dynamic speech motor assessment for Swedish-speaking children with childhood apraxia of speech, Logopedics Phoniatrics Vocology, 47:4, 230-238, DOI: [10.1080/14015439.2021.1943517](https://doi.org/10.1080/14015439.2021.1943517)

To link to this article: <https://doi.org/10.1080/14015439.2021.1943517>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 06 Jul 2021.



Submit your article to this journal [↗](#)



Article views: 2007



View related articles [↗](#)



View Crossmark data [↗](#)

RESEARCH ARTICLE



A preliminary validation of a dynamic speech motor assessment for Swedish-speaking children with childhood apraxia of speech

Susanne Rex^{a*} , Anders Sand^a , Edythe Strand^b, Kristina Hansson^c  and Anita McAllister^{a**} 

^aDepartment of Clinical Sciences, Intervention and Technology (CLINTEC), Division of Speech and Language Pathology, Karolinska Institutet, Stockholm, Sweden; ^bMayo Clinic College of Medicine, Rochester, MN, USA; ^cDepartment of Clinical Sciences, Lund, Logopedics, Phoniatrics and Audiology, Lund University, Lund, Sweden

ABSTRACT

Objective: Children with speech sound disorders (SSDs) have difficulties affecting different levels of speech production. For treatment to be beneficial, it is important to differentiate between Childhood Apraxia of Speech (CAS) – a motor speech disorder with deficits in speech praxis – and other SSDs (nonCAS-SSD). We have previously developed a motor speech examination Dynamisk motorisk talbedömning (DYMTA). We aimed to evaluate DYMTA's reliability and validity in a small-scale sample to estimate DYMTA's usability in diagnostic settings.

Methods: Speech, language, and oral motor abilities were assessed in 45 children between 40 and 106 months. Intra- and inter-rater reliability of DYMTA were analyzed. Further, DYMTA's ability to validly discriminate between children with CAS and nonCAS-SSD was assessed.

Results: The intra-rater reliability for the scores of DYMTA was strong, with ICCs ranging from 0.97 to 1.0. DYMTA total score had strong inter-rater reliability as evidenced both by the agreement estimates (DYMTA-A: 0.91 and DYMTA-B: 0.87) and the ICCs (0.97 and 0.96). Inter-rater reliability was also strong for the separate subscores on agreement estimates and for all subscores on ICCs, except for the Prosody subscores. DYMTA accurately discriminated between children with CAS and nonCAS-SSD in this small sample with an AUC of 0.92 for DYMTA-A and 0.94 for DYMTA-B.

Conclusions: With its focus on speech movements, DYMTA could serve as a valuable addition to other tests when assessing children's speech motor performance. This first examination suggests that DYMTA may be both a reliable and valid tool in the diagnostic process of SSD.

ARTICLE HISTORY

Received 12 February 2021

Revised 3 May 2021

Accepted 9 June 2021

KEYWORDS

Speech sound disorder;
differential diagnosis;
reliability; validity; DYMTA


Introduction

Children with speech sound disorders (SSDs) including childhood apraxia of speech (CAS) comprise a heterogeneous group. In the clinical setting, it may be a challenge to differentiate between CAS – a motor speech disorder with deficits in speech praxis – and other types of SSDs [1–3]. There are no biomarkers for any of the SSDs, therefore, diagnosis depends on careful observation of specific groups of agreed upon characteristics that are associated with a particular label for the subtype of SSD. Assessment tasks allow observations of speech across contexts, and dynamic assessment adds important information regarding severity and prognosis. Speech motor deficits present on a continuum of severity, and often involve linguistic deficits. A variety of assessment tasks including dynamic assessment is important to determine the level of impairment (i.e. motor or linguistic) or the relative contribution of the two [4]. The results of the assessment would then support differential diagnosis, provide evidence for severity and prognosis, and serve as indicator to the appropriate form of intervention [5]. For a

test to be used in the diagnostic process and treatment planning, it should be both reliable and valid [6]. Recently a motor speech examination in Swedish; Dynamisk motorisk talbedömning (DYMTA), was developed [7] and described [8]. In this study, we report on a first step toward estimating the reliability and validity for that test.

Dynamic motor speech assessment

In a dynamic motor speech assessment (DMSA) repetition of words of varying length and phonetic complexity is especially important in order to determine or rule out difficulties with motor speech planning and programming in children [9]. The clinician actively provides a cue, such as repeating more slowly or using a tactile cue and interacts with the child for each speech-item to facilitate accuracy and emerging skills. It is the child's response to those cues that are scored. Consequently, the DMSA provides information on movement accuracy, type of speech errors, and learning strategies/stimulability of speech movements.

CONTACT Susanne Rex  susanne.rex@ki.se  Division of speech and language pathology, CLINTEC, Karolinska Institutet, Sweden

*Department of Otorhinolaryngology, Speech and Language Pathology, Skåne University Hospital, Sweden.

**Medical Unit Speech and Language Pathology, Karolinska University Hospital, Sweden.

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

One such assessment is the Dynamic Evaluation of Motor Speech Skills (DEMSS) [5], a motor speech test designed to examine speech movements of children with severe speech deficits [4]. DEMSS motivated the design of a Swedish tool to assess performance associated with difficulties in motor speech programming (e.g. lengthened and disrupted coarticulatory transitions between sounds and syllables), but it specifically assesses articulatory accuracy, vowel accuracy, prosodic errors, and consistency of productions over repeated trials [10,11]. In order to reveal deficits underlying motor speech programming, dynamic assessment allows the clinician to observe characteristics not often seen in spontaneous speech or in static assessment such as traditional articulation tests [4,9].

The most commonly used discriminative features for CAS have been listed [12] and operationalized [13,14] and includes vowel error, consonant distortion, lexical stress error, syllable segregation, articulatory groping, intrusive schwa, voicing errors, slow rate, increased difficulties with multisyllabic words, resonance disturbance, difficulty achieving initial articulatory configuration, and occurrence of inconsistency over repeated trials. Some of the overt features will vary depending on the child's native language. Therefore, the phonotactics and prosodic elements of the Swedish language were considered when devising the test content [8]. Further, clinicians need to know how the characteristics of CAS may be represented in the language(s) of the child to be able to do a correct perceptual assessment.

Swedish phonology and prosody

Swedish has a complex prosodic system with contrasts of vowel length and quality, word stress (i.e. initial *versus* non-initial stress), as well as of tonal word accent (i.e. Accent 1; one-peaked fundamental frequency contour and Accent 2; two-peaked fundamental frequency contour) [15]. Stress affects the quality of both consonants and vowels in Swedish. The development of Swedish prosody is manifested in the child's speech as early as 18 months but continues to be refined through early childhood [16]. In a cohort of 25 Swedish children with delayed language development, 60% had difficulties producing the contrast of tonal word accent [17]. Results from both acoustic and perceptual evaluations verified this difficulty. Further, the acoustic analysis revealed overall flattened curves in the deviant F0 patterns indicating that the problem with tonal word accents was more a lack of producing the distinction between the accents than a substitution of one accent for the other, as previously proposed [18].

The Swedish language has 18 consonants distributed over four places of articulation (labial, dental/alveolar, palatal/velar, and glottal), five manners of articulation (stops /p, b, t, d, k, g/, nasals /m, n, ŋ/, fricatives /f, v, s, ʃ, ç, h/, liquids /l, r/, approximant /j/) and the distinction between voiced-voiceless stops, i.e. /b, d, g/ *versus* /p, t, k/. All consonants are established by the age of five, with the exception of /r/ that is acquired at six and regarding /s/ an interdental lisp may be evident in some 6-year-old children. The vowel

inventory displays an interplay between spectral dimensions (tongue height [F1], front-back tongue position [F2], and lip rounding or lip spreading) and temporal dimensions (vowel length in relation to consonant length) resulting in 18 vowels in most dialects [19]. In a previous study on typically developing Swedish children, we found vowels to be acquired by the age of three [8]. Only the high, over-rounded vowel /y:/ was developed somewhat later in a few children.

While DYMTA was motivated by the DEMSS, it was designed for a larger range of age and severity. Consequently, taking Swedish phonotactics, phonetics, and prosody into account, a second set of test items consisting of phonotactically more complex words was constructed, resulting in DYMTA-A and DYMTA-B [7,8].

DYMTA

DYMTA systematically varies phonotactic difficulty within targeted Swedish words and is administered using repetition with dynamic cueing to determine to what degree that cueing affects performance [20]. The procedure, scoring principles and the performance of typically developing children on the test have previously been described in detail [8]. DYMTA-A, the first part of DYMTA, was constructed with early developing syllable structures and phonology, to target younger children and/or those with more severe SSD. DYMTA-B was constructed to also target children with less severe SSD, and includes items with more complex transitional movements, which is more representative of everyday language. The selection of DYMTA-A *versus* DYMTA-B in the diagnostic process is dependent on the child's age and severity of the speech disorder.

DYMTA-A comprises 55 words with early developing phonemes and syllable shapes in eight subtests. All words are judged regarding *articulatory accuracy* (correct production of the target, with accurate smooth coarticulatory transitions between sounds and syllables), *vowel accuracy* (age-related vowel production), and *consistency* (consistency across repeated trials of the same word). Twenty-two of the words are also scored with respect to *prosody* (word stress and tonal accent). The results on these four aspects constitute the respective subscores and the sum of these subscores comprises the total score (with a maximum of 407). DYMTA-B subtests include all Swedish phonemes, syllable shapes, and prosodic aspects, thus increasing the motor programming demands on speech movements, relative to DYMTA-A. DYMTA-B has nine subtests, each focusing on a feature or spatial-temporal aspect associated with characteristics of CAS. These include voiced-voiceless and stop-fricative contrast, speech movement transitions, and consonant clusters. The Swedish prosodic aspects are also addressed with respect to word stress and tonal word accent. DYMTA-B comprises 71 words and utterances all targeted with respect to *consistency*, 62 words targeted for *articulatory accuracy*, 68 words for *vowel accuracy*, and 24 words scored with respect to *prosody*, with a maximum total score of 461. In Table 1, we list the subtests of DYMTA-A and

Table 1. Subtests and number of words/utterances in DYMTA-A and DYMTA-B, respectively.

DYMTA-A	Number of utterances	DYMTA-B	Number of utterances
1. CV	11	1. Simple syllables	13
2. VC	6	2. Voiced-voiceless contrast	6
3. CVCV	4	3. Dental-velar contrast	6
4. CVC	6	4. Stop-fricative contrast	6
5. CVC ₂	10	5. Consonant clusters	10
6. CV ₁ CV ₂	6	6. Word stress	8
7. C ₁ V ₁ C ₂ V ₂	6	7. Tonal word accent	6
8. Multisyllabic words	6	8. Multisyllabic words	10
		9. Increased length of utterance	6
	55		71

DYMTA-B, respectively. For a comprehensive overview of number of subscores and examples of words/utterances we refer to our previous study by Rex et al. [8].

A multidimensional scoring system (scores four to zero) is applied for *articulatory accuracy* to reflect the responsiveness to cuing, with higher scores indicating better performance and less dependency on cuing. A three-point scale is used for *vowel accuracy* (scores two to zero) while binary scoring is used for both *prosody* and *consistency*. DYMTA includes a manual, scoring protocols and a picture illustrating words from the two tests, to be used to elicit spontaneous speech [7].

DYMTA is the first test specifically aimed at motor speech assessment in Swedish-speaking children. Prior to DYMTA, only qualitative interpretations of performance for articulation and phonology were used to evaluate motor speech performance. There is a clinical need for a specific motor speech instrument to assist in the process of differential diagnosis of SSD. Psychometric evidence of reliability and validity, while important in test development and for a test to be used in diagnostic settings, is lacking in several assessment tools frequently used by speech-language pathologists (SLPs) for Swedish speaking children. Reliability estimates the degree of consistency of a measure and validity estimates how well a test measures what it claims to measure [6]. Types of reliability and validity and how they were calculated for DYMTA are further described in the method section.

Purpose

The purpose of this study was to provide a preliminary investigation of the reliability and validity of DYMTA. Specifically, we evaluated DYMTA's intra- and inter-rater reliability and its ability to validly discriminate between children with SSD who exhibit CAS from those who do not.

Method

Participants

Participants included 45 children (31 males and 14 females) between 40 and 106 months ($M=68$ months, $SD=16$ months, $Mdn=62$ months), who were consecutively recruited by the first author in a regional clinical setting at

the Speech-Language Pathology Clinic, ENT Department, Skåne University Hospital, during a defined period of 1½ years. The inclusion criteria for participation were (a) a referral question of SSD and/or an explicit question about suspected CAS (there is a referral bias in our clinic such that more children with suspected CAS are likely to be referred than if referrals were completely evenly distributed between clinics; as a consequence of this bias, quite a large proportion of children in our sample were diagnosed with CAS) (b) age between 3 and 9 years old and (c) normal or adjusted-to-normal hearing and (d) Swedish as first language (no children were learning a second language). Exclusionary criteria included structural deficits (e.g. cleft palate), autism spectrum disorder (ASD), and lack of participation in the DYMTA assessment. During the recruitment period, 50 children met the criteria and were asked to participate, 49 accepted, 2 were found to have ASD, 2 discontinued, leaving a study group consisting of 45 children. Descriptive data for all participants are shown in Table 2.

Procedure

All participants completed a standardized assessment battery examining speech-, language- and oral motor performance administered by the first author. A test of phonology, Linköpingsundersökningen (LINUS) [21], was administered to the children through picture naming. The short version of LINUS consisting of 40 words was used. Performance was used to obtain phonetic and phonemic inventories, identify consistent developmental substitution errors and quantified as percent consonants correct (PCC) [22], percent vowels correct (PVC), and percent words correct (PWC) based on narrow transcription.

Oral motor performance was assessed with the Verbal Motor Performance Assessment for Children (VMPAC) [23]. For this study, the first 37 items of VMPAC were summarized as a total non-verbal oral motor score (OMS) comprising orofacial integrity and non-speech oromotor skills. Normal structures were found in all children except one child having a slight facial asymmetry. All had normal reflexes, normal trunk and neck stability and respiratory coordination.

Language production was assessed with GRAMBA (Grammatiktest för barn) [24] a Swedish test assessing different noun and verb forms as well as word order with norms for children 3–5;11. Raw GRAMBA scores are shown in Table 2 for all children. To assess language comprehension, the Swedish version of TROG-2 (Test for Reception of Grammar – Second Edition) [25] was used from 4 years and Nya SIT (Språkligt Impressivt Test för barn) [26] for the younger children. Swedish TROG-2 has established norm data from 4 years. TROG-2 results were transformed to percentiles according to the manual. The SIT manual presents reference data in raw scores indicating average number of errors per age group. Performance on SIT is shown in number of errors and marked with * in Table 2.

Table 2. Demographics and speech- language data for individual participants ordered by DYMTA-A scores within diagnostic groups..

Group	ID	DYMTA-A Total score	DYMTA-B Total score	Sex	Age (months)	PCC	PVC	PWC	GRAMBA	Receptive	OMS
SSD (non-CAS)	24	407	445	M	103	98	100	98	41	53 ^b	87
	16	407	408	M	57	89	97	72	43	94	97
	34	407	405	F	63	89	100	55	39	86	100
	6	406	433	M	61	86	99	58	44	98	100
	15	397	343	M	50	71	95	32	39	91	91
	10	395	373	M	59	62	91	25	41	97	98
	38	386	446	M	74	98	97	90	41	87	100
	14	381	404	M	58	79	96	40	44	99	96
	18	369	395	M	66	77	99	40	42	82	97
	8	361	335	F	50	63	98	25	43	63	94
	4	357	334	F	52	64	94	20	29	91	96
	18	340	311	M	59	65	92	28	39	45	89
	33	391	367	M	58	81	97	42	42	99	95
	35	381	340	M	58	62	88	30	— ^a	15 ^{*a}	77
	17	365	401	M	92	86	99	65	43	58	99
mild CAS	1	361	354	F	57	74	92	38	43	99	95
	20	331	318	M	81	70	97	32	44	75	100
	7	384	311	F	98	55	95	20	44	66	99
	9	375	323	M	91	75	95	28	42	21	99
	36	364	290	M	62	68	100	22	41	66	92
CAS	19	359	318	F	88	59	95	20	40	70	98
	29	357	314	M	62	62	90	15	16 ^a	10	90
	23	348	341	F	106	77	95	52	38 ^a	39	95
	5	342	288	M	81	64	86	15	41	84	96
	39	333	299	F	61	59	96	25	42	18	79
	45	329	252	F	77	42	91	15	22 ^a	3 ^a	95
	2	321	313	M	74	48	92	10	42	93	96
	11	312	289	F	90	54	96	12	44	32	100
	21	295	267	M	81	49	96	8	44	84	99
	41	295	262	M	82	72	85	30	43	63	97
	13	289	267	M	64	48	95	10	38	96	98
	3	281	306	M	53	82	88	40	44	99	98
	12	280	269	F	90	49	96	15	44	45	98
	44	279	257	F	76	60	96	18	26 ^a	0 ^a	88
	37	274	225	M	61	34	94	10	37	86	100
	25	272	270	M	60	56	90	15	36	66	100
	40	269	190	M	68	42	94	5	34	95	97
	32	258	268	M	69	57	88	10	23 ^a	73	100
	26	256	273	M	54	56	92	10	41	58	74
	31	221	199	M	46	27	65	0	12 ^a	8 [*]	92
	28	205	181	M	56	24	84	0	38	99	100
	30	186	129	M	45	32	85	5	24	10 [*]	76
	42	165	110	F	67	31	62	2	36	61	93
	22	164	124	F	40	24	67	0	— ^a	17 [*]	71
	43	153	115	M	44	14	64	0	2 ^a	14 [*]	71

ID: identification; PCC: percent consonant correct; PVC: percent vowels correct; PWC: percent words correct (PCC, PVC, and PWC are calculated from LINUS phonology test); Receptive = Test for Reception of Grammar (TROG2) for 4 years old and older (percentiles) and Språkligt Impressivt Test för barn (SIT) for 3-year-olds, with one exception (mean number of errors for 36–47 months is 14 and for 48–59 months 8 errors, of maximum score 46), *indicates SIT scores; GRAMBA: Grammatiktest för Barn (total score, maximum 44); OMS: oral motor score derived from VMPAC (from subtests *focal oromotor control* and *sequencing maintenance control*). F: female, M: male. Language scores followed by ^aare considered in the clinically impaired range, ^bthis child has word retrieval difficulties not captured in the GRAMBA score. Em dashes indicate missing values due to lack of child attention that prevented completion of the whole test.

This study was conducted in the context of a clinical practice at the Skåne University Hospital, Sweden. DYMTA, which focuses on observations related to movement accuracy and prosody (reflecting speech motor planning) *versus* phonologic development, was administered as part of the assessment battery. Reports of speech diagnosis were completed immediately after the assessment session. The reports included a statement of differential diagnosis based on the examiner's clinical observations and examination of language and speech test results. Observations were noted from the online assessment of DYMTA as well as all other assessment tasks administered. In order to reduce potential bias, the DYMTA scoring was processed and calculated more than a year after the actual administration of the test. In the

diagnostic process, in our clinical setting, a CAS diagnosis is reported for those children whose speech characteristics are consistent with that diagnosis. However, CAS occurs on a continuum of severity and some children with phonologic impairment and/or residual phonetic errors also have symptoms indicating deficits in speech motor planning/programming (e.g. occasional vowel distortions; mild or infrequent syllable segmentation; occasional prosodic errors). This continuum was represented as CAS, mild CAS, and other SSDs (in this study referred to as nonCAS-SSD).

Approval for the study was given from The Regional Ethical Review Board in Lund (Dnr: 2013/24). All data were handled according to data protecting rules. Informed written consent was obtained from parents/legal guardians of the

participants. In addition, the children gave their assent to participate in the study just before assessment started.

Reliability analyses

We analyzed intra- and inter-rater reliability of DYMTA. Nine randomly selected children (20%) were used for the analysis. Both children with CAS and nonCAS-SSD were represented in the sample (six children with CAS and three nonCAS-SSD). Since reliability may be more difficult to obtain when there is a more severe impairment, we examined our randomly selected sample and found that the severity continuum was represented with four severe, one moderate, and one mild CAS. To evaluate intra-rater reliability, the randomly selected recordings were rated again by the first author. The second rating was blinded and occurred almost a year after the first rating. We analyzed percentage of agreement between the two timepoints point-by-point for each word and report this for the total score and the subscores of DYMTA. We also analyzed the inter-class correlation coefficients (ICCs) for DYMTA total score and subscores, see below for more information. For inter-rater reliability, the nine randomly selected recordings were rated again by another speech pathologist with extensive experience of SSDs. The first and second rater each scored nine administrations of DYMTA-A and DYMTA-B from video/audio-recordings of the original assessment. The second rater had attended a course in administering DYMTA and was routinely using DYMTA in the clinical setting. We analyzed percentage of agreement between the two raters, “point-by-point,” for each word and report this for the total score and the subscores of DYMTA. We also analyzed the ICCs for DYMTA total score and subscores and report reliability calculated for a single measurement when raters are randomly sampled from a larger population of raters (ICC [2,1]) [27]. The ICC was calculated by using a random-intercept mixed-effect linear model, with participants as random effects and only an intercept as the fixed effect, following Strand et al. [4]. In this model, the ICC is the proportion of variance that can be attributed to participant-to-participant variability and $1 - \text{ICC}$ can be thought of as the proportion of variance attributed to rater-to-rater variability (or between timepoints for intra-rater reliability).

Validity analysis

We assessed DYMTAs ability to discriminate between children with CAS (CAS and mild CAS) and those without CAS based on their clinical diagnosis. With a perfect test, children with and without a defined disability would get non-overlapping test scores and discriminating between groups would be easy. With more realistic tests, there will be an overlap in test scores. A researcher can specify a specific cutoff score that maximizes *sensitivity* – correctly classifying children with CAS as having CAS – or *specificity* – correctly classifying children without CAS as not having CAS – or a cutoff that balances the two. By plotting

sensitivity and specificity pairs for various cutoffs in the test scores obtained here, we created a receiver operating characteristic (ROC) curve of DYMTA (see panel B and D in Figure 1 for examples). From this, we report DYMTA’s area under the receiver operating characteristic curve (AUC). The AUC is a non-parametric measure of how well a test discriminates the diagnostic groups (the probability of correct classification). We report AUC, with bootstrapped 95% confidence intervals (CI; 10’000 resamples), for DYMTA total score and the separate subscores. Sensitivity, specificity, and likelihood ratios are reported for a cutoff that maximizes the positive likelihood ratio, as well as report the specificity for a cutoff with 90% sensitivity (following Strand et al. [4]). The *positive likelihood ratio* is how much more likely a child with CAS is identified by the test (DYMTA), than a child without CAS is falsely identified. The *negative likelihood ratio* is, how much more likely the test (DYMTA) correctly identifies a child as not exhibiting CAS, than misses a child with CAS.

Results

Reliability

Table 3 details the point-by-point intra- and inter-rater agreement and ICCs for the total scores and separate subscores of DYMTA-A and DYMTA-B. The intra-rater results indicate very strong intra-rater reliability for both DYMTA-A and DYMTA-B total score and separate subscores. Both DYMTA-A and DYMTA-B total score also had strong inter-rater reliability as evidenced both by the agreement estimates (0.91 and 0.87) and the ICCs (0.97 and 0.96). Inter-rater reliability was also strong for the separate subscores on agreement estimates and for all subscores on ICCs, except for Prosody. On the Prosody subscore for DYMTA-A, most of the nine children randomly selected for the inter-rater reliability estimation had high scores, but one rater assessed two children as having somewhat fewer correct productions. This small overall variability, and the inter-rater difference for these two children’s scores, produced the low ICC values for prosody, see Table 3. A similar scenario was also true for the Prosody subscore in DYMTA-B.

Validity

Children diagnosed with CAS and nonCAS-SSD were fairly well separated on DYMTA-A total score (Figure 1 panel A) and even more so on the DYMTA-B total score (panel C). Panel B and D in Figure 1 illustrate the estimated ROC-curve based on DYMTA-A total score and DYMTA-B total score, respectively, and Table 4 provides AUC and other measures of how well DYMTA-A and -B discriminated among children with CAS or nonCAS-SSD. In general, DYMTA could accurately discriminate between children with SSD and CAS (i.e. AUC 0.92 for DYMTA-A and 0.94 for DYMTA-B total test score). Two potential cutoffs in DYMTA-A total score are illustrated in Figure 1, panel A.

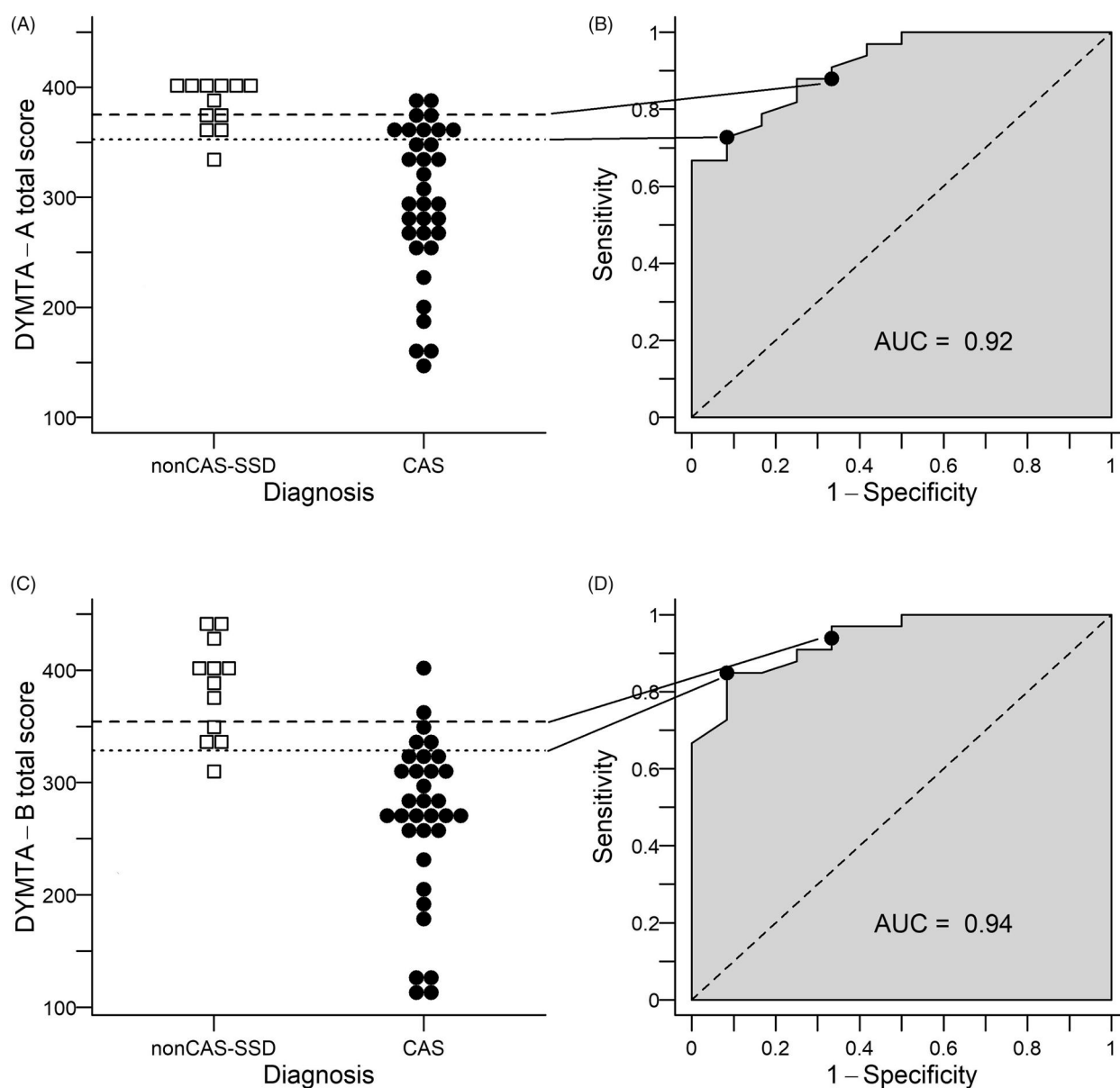


Figure 1. Panel A and C illustrates performance on DYMTA-A total score and DYMTA-B total score, respectively, separately for children with CAS and nonCAS-SSD. In these figures, each individual is shown as a separate point. Panel B and D illustrates the ROC-curve for the DYMTA-A and DYMTA-B. Two potential cutoffs are illustrated in panel A, one based on the binomial logistic regression (dashed line) and one based on the positive likelihood ratio (dotted line) and the sensitivity and specificity pair related to these cutoffs are illustrated in panel B. Similarly, for panel C and D. The AUC is shown in panel B and D as the shaded area.

Table 3. Estimates of reliability for DYMTA-A and DYMTA-B total scores and subscores.

Measure	Intra-rater reliability		Inter-rater reliability	
	Point-by-point agreement mean (range)	ICCs estimate [CI]	Point-by-point agreement mean (range)	ICCs estimate [CI]
DYMTA-A				
Total score	0.98 (0.94–1.00)	0.99 [0.99, 1.00]	0.91 (0.84–0.97)	0.97 [0.88, 0.99]
Vowel accuracy	0.98 (0.91–1.00)	0.98 [0.93, 0.99]	0.92 (0.84–1.00)	0.82 [0.48, 0.94]
Artic accuracy	0.97 (0.87–1.00)	0.99 [0.99, 1.00]	0.91 (0.84–0.98)	0.98 [0.92, 0.99]
Prosody	0.99 (0.95–1.00)	0.88 [0.66, 0.96]	0.89 (0.73–1.00)	0.03 [–0.53, 0.57]
Consistency	0.98 (0.95–1.00)	0.99 [0.97, 1.00]	0.91 (0.82–0.98)	0.96 [0.88, 0.99]
DYMTA-B				
Total score	0.95 (0.91–0.98)	0.99 [0.97, 1.00]	0.87 (0.79–0.91)	0.96 [0.89, 0.99]
Vowel accuracy	0.98 (0.92–1.00)	0.97 [0.92, 0.99]	0.90 (0.75–0.97)	0.49 [–0.04, 0.81]
Artic accuracy	0.95 (0.92–0.98)	0.99 [0.99, 1.00]	0.89 (0.78–0.98)	0.99 [0.97, 1.00]
Prosody	0.96 (0.75–1.00)	0.73 [0.35, 0.91]	0.83 (0.75–0.96)	0.0 [–0.55, 0.55]
Consistency	0.92 (0.82–0.96)	0.98 [0.90, 0.99]	0.83 (0.66–0.88)	0.92 [0.77, 0.98]

ICC: intra-class correlation coefficient; CI: 95% Confidence interval.

Table 4. Estimates of discrimination ability of DYMTA-A and DYMTA-B total scores and subscores.

Measure	AUC (95% CI)	Cutoff based on LR+	LR+/LR- (based on LR+)	Sensitivity/specificity based on LR+	Specificity at 90% sensitivity
DYMTA-A					
Total score	0.92 [0.82, 0.98]	352.5	8.73/0.30	0.73/0.92	0.67
Vowel accuracy	0.89 [0.80, 0.97]	105.5	8.36/0.33	0.70/0.92	0.75
Artic accuracy	0.90 [0.80, 0.98]	173.0	8.36/0.33	0.92/0.67	0.91
Prosody	0.60 [0.47, 0.74]	20.5	3.64/0.76	0.30/0.92	–
Consistency	0.91 [0.81, 0.98]	47.5	8.73/0.30	0.73/0.92	0.50
DYMTA-B					
Total score	0.94 [0.86, 0.99]	328.5	10.18/0.17	0.85/0.92	0.67
Vowel accuracy	0.89 [0.79, 0.97]	125.0	9.45/0.23	0.79/0.92	0.42
Artic accuracy	0.92 [0.82, 0.98]	125.5	9.09/0.26	0.76/0.92	0.67
Prosody	0.73 [0.56, 0.88]	19.5	4.73/0.66	0.39/0.92	0.50
Consistency	0.86 [0.71, 0.97]	55.5	4.73/0.25	0.79/0.83	0.58

AUC: area under the receiver operating characteristic curve; LR+: likelihood ratio for a positive test result; LR–: likelihood ratio for a negative test result. For Prosody subscore in DYMTA-A no reliable cutoff that produced 90 % sensitivity could be produced and thus specificity at such a cutoff could not be estimated.

In panel B, the corresponding pair of sensitivity and specificity values are shown. One cutoff (dashed line) is the minimal value for which the probability of having CAS is greater than 50%, this was empirically derived from a binary logistic regression model. The other cutoff (dotted line) is the cutoff that maximizes the positive likelihood ratio (i.e. the cutoff that maximizes our certainty that a positive classification is correct). At these cutoffs, the diagnostic odds ratio for DYMTA-A was 29.3 (95% CI [3.3, 261.0]) and for DYMTA-B it was 61.6 (95% CI [6.4, 588.9]). Similarly, panel C and D illustrates DYMTA-B total score and its corresponding ROC curve.

Discussion

The aim of this study was to evaluate the reliability and validity of DYMTA. Generally, we estimated DYMTA-A and DYMTA-B to have similar reliability and validity values and thus discuss these combined as DYMTA, unless otherwise noted.

Reliability

For DYMTA to be used in clinical settings, it is important that it has both acceptable intra- and inter-rater reliability. DYMTAs strong intra-rater reliability (98% and 96% mean agreement for DYMTA-A and DYMTA-B, respectively) suggests that DYMTA scoring is not heavily impacted by day-to-day measurement error on the clinician's part. This is particularly important in dynamic assessment, including multidimensional scoring, which is putting a lot of pressure both on the child and the clinician. Based on the strong agreement between the two raters tested here (91% and 87% inter-rater agreement for DYMTA-A and DYMTA-B, respectively), DYMTA-scoring does not seem to be heavily impacted by which clinician is performing the scoring. However, we note that this result may not generalize to clinicians without proper training using the test. This is a potential risk in all clinical assessments and especially when test takers are young children with possible additional difficulties with attention, temperament, or language comprehension [28].

The reliability for individual DYMTA subscores were also examined demonstrating the intra-rater mean agreement as strong for all subscores. For inter-rater reliability, the percent agreement and ICC corroborated strong reliability estimates for most subscores. However, although the percent agreement was high for the *Prosody* subscore, the ICC approach did not generate robust estimates of inter-rater reliability. Several factors may have led to the less robust ICC estimate for the *Prosody* subscores. First, there are fewer *Prosody* items, 22 and 24, respectively, for DYMTA-A/DYMTA-B, compared to 55/62 for articulatory accuracy, 55/68 for vowel accuracy, and 55/71 for consistency. Second, variation was limited, and a ceiling effect occurred among the randomly selected participants used for evaluating inter-rater reliability. Third, it could simply be that assessing prosody is more challenging and something SLPs are less trained to do, compared to other speech problems. Finally, Swedish prosody is a complex system with several stress patterns and two tonal word accents, where dialects and regional variations in accentuation and intonation add to the complexity [15]. Consequently, listeners may have different perceptual references when differentiating between word tones. In a study on prosodic difficulties in Swedish children with speech- and language disorder, prosodic features were tested in 12 subtests and the tonal word accent sub-test had the lowest inter-rater agreement [18]. Also, the flattened tonal accent curve in many children with speech- and language disorder found by Samuelsson and Löfqvist [17] could affect the perceptual rating or agreement, where one rater interprets this “in-between” accentuation as Accent 1 and the other as Accent 2.

These findings suggest that perceptual training for prosody should be included in graduate speech pathology programs, as well as in continuing education offerings. We are planning for training *via* a student e-learning platform, and also to make a training video to be included with the test. Furthermore, an easily accessible and user-friendly software for acoustical assessment of prosody could also be a solution. In a potential revision of DYMTA we will look at the selected items and add targets for prosodic scoring, preferably sentences, to be sure prosody is assessed in the most reliable way. Despite the somewhat challenging perceptual judgment of Swedish prosody, we conclude that inter-rater

reliability was strong for all DYMTA subscores based on high percent agreement.

Validity

If DYMTA is to be used as a clinical tool to identify children with CAS, it is important that it does so with high accuracy. Overall, the total score of DYMTA seems to validly separate children with CAS and nonCAS-SSD (AUC = 0.92 and 0.94). Further, the specific cutoff we determined, using the positive likelihood ratio, had acceptable sensitivity (0.73 and 0.85) and specificity (0.92), and a high diagnostic odds ratio (29 and 62), demonstrating that DYMTA can be a valuable tool in clinical evaluations. To put these estimates into perspective, DYMTA seems to discriminate on a similar level as did DEMSS in Strand et al.'s evaluation [4]; DEMSS had an AUC of 0.93 and 65% sensitivity and 97% specificity. Note that differences in sample due to participant characteristics, and exact derivation of thresholds should temper too fine-grained direct comparisons. Although separate subscores of DYMTA should not be used individually to diagnose children with CAS, our results suggest that they are meaningful parts of DYMTA to the extent that they separate children with CAS from children with other speech deficits. Based on the sample used, each subscore, apart from the *Prosody* subscore, had acceptable AUC values.

It should be noted that sensitivity, specificity, as well as positive and negative likelihood ratios, are not properties of a test *per se*, but rather properties of a test used in a specific sample. That is, the “base rate,” or prevalence, of affected individuals in a sample will influence these statistics. In our sample, for example, we have more children with CAS than nonCAS-SSD and a consequence of this may be an underestimation of some statistics describing the discrimination ability of DYMTA (i.e. the sensitivity and specificity). The AUC, on the other hand, is unaffected by the base rate in the sample (except for robustness of estimation, of course) and thus is the take-home evaluation of DYMTA discriminative ability.

Clinical implications

Our present evaluation suggests that DYMTA can positively contribute to clinical decisions regarding the likelihood of a child having CAS. Lower DYMTA scores represent higher risk of CAS with scores below 353 for DYMTA-A and 329 for DYMTA-B indicating a diagnosis of CAS. These cutoffs are based on maximizing certainty that a positive classification is correct. However, it is important to remember that DYMTA is one component in a more comprehensive diagnostic assessment of speech and language. Because we seek to determine how linguistic *versus* motor deficits contribute to the speech disorder, we need to know about phonological errors besides speech motor skills. Additionally, the DYMTA scoring system requires knowledge of the child's phonology and, therefore, a phonology test should be administered before DYMTA. For example, if a 4-year-old *versus* an 8-year-old child would substitute /r/ with /j/, this

would be reflected differently in the scoring. Likewise, the language ability of the child may influence the results on DYMTA. For example, if a child has attention and/or language comprehension deficits it will be possible to help by adjusting the instructions, accordingly, provided that the results from the language testing are known in advance. This is to make sure to get the best effort and performance from the child. Further, oral motor competence is an important part of the information incorporated in the diagnostic assessment. The structural-functional examination will identify or rule out the presence of anatomical deficits, oral motor function (the integrity of cranial nerves V, VII, IX, X, XI, XII) and oral dyspraxia. Thorough information on all co-existing deficits is important to the clinician in the diagnostic procedure, and it will also be valuable to consider during intervention planning.

Limitations

This study is a preliminary evaluation of DYMTA's diagnostic value. The main limitation here is the small-scale sample and especially that only twelve children with nonCAS-SSD were included (of a total of 45 participants). The small proportion of children with nonCAS-SSD was probably a consequence of the referral bias in our clinic (see participants above). The non-representative ratio of children with nonCAS-SSD *versus* CAS in our sample makes it difficult to fully evaluate the positive predictive value of DYMTA, as such an estimation would be biased without taking a correct estimation of the prevalence of CAS into account (but note, again, that AUC is unaffected by this). Further, with regards to intra- and inter-rater reliability, in a DMSA the clinician is not only *scoring* the test, but also dynamically *cueing* the child. Both *scoring* and *cueing* during assessment can vary between days (*vis-à-vis* intra-rater reliability) and between clinicians (*vis-à-vis* inter-rater reliability). Because we estimated reliability based on rescoring video recorded assessments, we have not estimated how much measurement error there may be in the *cueing* part of DYMTA.

Conclusions

DYMTA has been developed to facilitate differential diagnosis of children with characteristics of CAS. This article reports preliminary data showing promise for good reliability and validity for this purpose. It is important to recognize that DYMTA is not intended to replace any other test but will, with its focus on speech movements and speech motor control, serve as a valuable complement to other tests, when assessing children's speech performance.

Acknowledgments

We want to send our warmest appreciation to the participating children and their families.

Disclosure statement

Rex, McAllister, and Hansson, are authors of the Swedish DYMTA manual and have received royalties from sales. The five authors alone are responsible for the content and writing of the paper.

Notes on contributors

Susanne Rex is a Speech-Language Pathologist with a clinical position at Skåne University Hospital, Sweden, parallel PhD studies at the Division of Speech and Language Pathology, CLINTEC, Karolinska Institutet, Stockholm. She is also a clinical lecturer at Lund University. Her main research interest is speech and language in children with speech disorders and methodology for assessment of suspected CAS.

Anders Sand is a post doc at the Division of Speech and Language Pathology, CLINTEC, Karolinska Institutet. His main research interest is to focus methodology and statistical inference in speech and language pathology on practical and clinical relevance.

Edythe Strand is a Professor Emeritus, Mayo College of Medicine and former Division Head, Speech Pathology, Department of Neurology, Mayo Clinic, Rochester, Minnesota. She is currently an affiliate professor in the Department of speech and Hearing Sciences at the University of Washington, Seattle. She frequently gives lectures on the assessment and treatment of motor speech disorders in children and adults, management of communication disorders in degenerative neurologic disease, and neuroanatomy. Her primary clinical and research interests include assessment and treatment of children and adults with neurologic speech and language disorders.

Kristina Hansson is Associate Professor in Speech-Language Therapy at Lund University, Sweden. Her research interest is language development and language difficulties in children with developmental language disorder and children with hearing impairment, with special focus on grammar, semantics, and interaction.

Anita McAllister is Associate Professor and Speech-Language Pathologist at the Division of Speech and Language Pathology, CLINTEC, Karolinska Institutet and at the Medical Unit Speech and Language Pathology, Karolinska University Hospital. Her main research interests include assessment and effects of oral sensorimotor and speech disorders on speech production and effects of intervention, and voice disorders and communication related to environmental factors such as soundscape. Both areas focus on children.

ORCID

Susanne Rex  <http://orcid.org/0000-0002-8192-0065>
 Anders Sand  <http://orcid.org/0000-0002-6626-5197>
 Kristina Hansson  <http://orcid.org/0000-0002-0467-5190>
 Anita McAllister  <http://orcid.org/0000-0003-2208-0630>

References

- [1] Dodd B. Differential diagnosis of pediatric speech sound disorder. *Curr Dev Disord Rep*. 2014;1(3):189–196.
- [2] Lewis BA, Avrich AA, Freebairn LA, et al. Subtyping children with speech sound disorders by endophenotypes. *Top Lang Disord*. 2011;31(2):112–127.
- [3] Shriberg LD, Fourakis M, Hall SD, et al. Extensions to the speech disorders classification system (SDCS). *Clin Linguist Phon*. 2010;24(10):795–824.
- [4] Strand EA, McCauley RJ, Weigand SD, et al. A motor speech assessment for children with severe speech disorders: reliability and validity evidence. *J Speech Lang Hear Res*. 2013;56(2):505–520.
- [5] Strand EA, McCauley RJ. *Dynamic evaluation of motor speech skill (DEMSS) manual*. Baltimore (MD): Brookes Publishing; 2019.
- [6] Dollaghan CA. *The handbook for evidence-based practice in communication disorders*. Baltimore (MD): Paul H Brookes Publishing Co, Inc.; 2007.
- [7] Rex S, McAllister A, Hansson K. *Dynamisk motorisk talbedömning (DYMTA) manual*. Lund, Sweden: Kunskapsutveckling i Lund AB; 2016. ISBN: 978-91-976067-2-1.
- [8] Rex S, Hansson K, Strand EA, et al. Performance of Swedish children on a dynamic motor speech assessment. *Int J Speech-Lang Pathol*. 2021. DOI:10.1080/17549507.2020.186230024.
- [9] Strand EA. Appraising Apraxia: when a speech-sound disorder is severe, how do you know if it's childhood apraxia of speech? *Leader*. 2017;22(3):50–58.
- [10] Terband H, Maassen B, Maas E. A psycholinguistic framework for diagnosis and treatment planning of developmental speech disorders. *Folia Phoniatr Logop*. 2019;71(5–6):216–227.
- [11] ASHA. *Childhood Apraxia of Speech: nomenclature, definition, roles and responsibilities, and a call for action [Position statement]*. Rockville (MD): Author; 2007.
- [12] Shriberg LD, Lohmeier HL, Strand EA, et al. Encoding, memory, and transcoding deficits in Childhood Apraxia of Speech. *Clin Linguist Phon*. 2012;26(5):445–482.
- [13] Iuzzini-Seigel J, Hogan TP, Guarino AJ, et al. Reliance on auditory feedback in children with childhood apraxia of speech. *J Commun Disord*. 2015;54:32–42.
- [14] Iuzzini-Seigel J, Murray E. Speech assessment in children with childhood apraxia of speech. *Perspect ASHA Sigs*. 2017;2(2):47–60.
- [15] Bruce G. Components of a prosodic typology of Swedish intonation. In: Riad T, Gussenhoven C, editors. *Tones and tunes*. Berlin, Germany: Mouton de Gruyter; 2007. p. 113–146.
- [16] Engstrand O, Williams K, Lacerda F. Does babbling sound native? Listener responses to vocalizations produced by Swedish and American 12- and 18-month-olds. *Phonetica*. 2003;60(1):17–44.
- [17] Samuelsson C, Löfqvist A. The role of Swedish tonal word accents in children with language impairment. *Clin Linguist Phon*. 2006;20(4):231–248.
- [18] Samuelsson C, Nettelbladt U. Prosodic problems in Swedish children with language impairment: towards a classification of subgroups. *Int J Lang Commun Disord*. 2004;39(3):325–344.
- [19] McAllister R, Flege JE, Piske T. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *J Phonet*. 2002;30(2):229–258.
- [20] Lidz CS, Peña ED. Dynamic assessment: the model, its relevance as a nonbiased approach, and its application to Latino American preschool children. *LSHSS*. 1996;27(4):367–372.
- [21] Blumenthal C, Hammarström I. Linus – L. Linköpingsundersökningen. – [in Swedish]. A phonological test for children from 3 years of age. 2014. Available from: <http://liu.divaportal>
- [22] Shriberg LD, Kwiatkowski J. Phonological disorders III: a procedure for assessing severity of involvement. *J Speech Hear Disord*. 1982;47(3):256–270.
- [23] Hayden A, Square P. *VMPAC, verbal motor production assessment for children*. San Antonio (TX): The Psychological Corporation; 1999.
- [24] Hansson K, Nettelbladt U. *GRAMBA - Grammatiktest för barn*. Lund, Sweden: Studentlitteratur; 2010. ISBN: 9789144069296.
- [25] Bishop D. *TROG-2 test for reception of grammar*. 2nd ed. London: Pearson; 2009.
- [26] Hellquist B. *Nya SIT—Språkligt Impressivt Test för barn*. Lund, Sweden: Studentlitteratur; 2011. ISBN: 9789144078243.
- [27] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428.
- [28] McCauley RJ, Strand EA. A review of standardized tests of nonverbal oral and speech motor performance in children. *Am J Speech Lang Pathol*. 2008;17(1):81–91.