

# Italian Validation of a Test to Assess Dysarthria in Neurologic Patients: A Cross-Sectional Pilot Study

De Biagi Francesca<sup>1</sup>, Frigo Anna Chiara<sup>2</sup>, Turolla Andrea<sup>1</sup>, Nordio Sara<sup>1\*</sup>, Berta Giulia<sup>1</sup> and Meneghello Francesca<sup>1</sup>

<sup>1</sup>IRCCS Foundation Hospital San Camillo, Via Alberoni, 70, postcode 30126, Venice, Italy

<sup>2</sup>Dipartimento di Scienze Cardiologiche, Toraciche e Vascolari, University of Padua, Padua, Italy

## Abstract

Dysarthria is a motor speech disorder that results from an impairment of the muscles devoted to speech production, thus affecting the movements of the orofacial district. The type and severity of dysarthria depend on which structures of the central or peripheral nervous system are affected. Due to the vast range of acute and progressive neurological disorders that may cause dysarthria, its prevalence may be not negligible. In Italian clinical practice dysarthria is assessed using a standardized protocol, which has never been validated. The aim of the study is to explore the intra- and inter- reliability of a short-form of a protocol to assess dysarthria and compare scoring of the test face-to-face versus via video of patient assessment, which is broadly used in the Italian clinical practice.

50 dysarthric patients were enrolled for this pilot study and assessed by "Protocollo di Valutazione della Disartria". Lin's Concordance Correlation Coefficient (CCC) determined the consistency of measurements between the same rater and among different raters with different levels of expertise. Scoring was done both during face-to-face assessments and while watching video recordings of patients' evaluations.

Results indicated a good consistency of ratings in repeated measures over time (video-assessment intra-rater CCC>0.8). Nevertheless, inter-rater reliability was less satisfactory (video-assessment scoring inter-rater CCC<0.8), especially in the face-to-face administration of the protocol (face-to-face/video inter-rater CCC<0.8).

In conclusion, the protocol showed to have potential clinical utility to assess dysarthria in neurological patients, due to its completeness and ease of training and administration. However generalizations of the findings are limited, due to the characteristics of the study. Indeed, further research is required for a better validation of the instrument.

**Keywords:** Dysarthria; Assessment; Validation

## Introduction

The term dysarthria refers to an altered speech production resulting from a neurological injury involving the motor component of speech process.

Dysarthria's aetiology may be associated with both degenerative disorders and acute illnesses [1,2]. Even though it is difficult to appraise the exact prevalence and incidence of dysarthria within the general population, the disorder is not rare [1]. In fact, it has been estimated that dysarthria may account for 54% of all acquired communication neurogenic disorders [3].

Dysarthria may have a severe impact on participation and social interactions [4] and consequently on the quality of life of patients. It has been highlighted that even mild dysarthria may have significant social and psychological effects [5]. For these reasons, much research in recent years has focused attention in developing protocols that assess functioning and level of social participation, rather than impairment [6].

Generally, the assessment should determine the presence, nature and extent of impairment to support the differential diagnosis towards specific management [7]. Moreover, it should identify the primary problems in order to plan the proper goals of treatment [8].

Assessment procedures should comprise an evaluation of medical history, examination of speech structures, perceptual analysis of speech and judgment of intelligibility [9].

Assessment of intelligibility is crucial to set the right goals of treatment and it should be the main outcome measurement in all the cases of speech disorders. However, conflicting recommendations exist on how to measure it [10].

Many impairment-based protocols have been developed to assess dysarthria. Previous studies in Ireland<sup>6</sup> and UK<sup>5</sup> have reported that the mostly used tools are the Frenchay Dysarthria Assessment [11] and the Robertson Profile [12]. However, to our knowledge, the only tool that allows perceptual analysis of speech available in Italian is "Profilo di Valutazione della Disartria" [1], the normative data for which have been provided using a cross-cultural adaptation of the Robertson Profile. This tool is divided into eight subscales (i.e., respiration, voice, facial musculature, diadochokinesis, reflex, articulation, intelligibility, prosody), each one including several items. Each item has a score ranging from 1 (worst) to 4 (best). The internal construct validity was investigated through a Rasch analysis in a sample of 196 patients [13]. The results of the study suggested the possibility of creating a short version of the test with a rescoring of the items into a three-point scale.

## Aim

The main aim of the current study is to measure the reliability of a modified tool for the assessment of speech impairments (i.e., dysarthria)

**\*Corresponding author:** Nordio Sara, IRCCS Foundation Hospital, San Camillo, Via Alberoni, 70, postcode 30126, Venice, Italy, Tel: 0039 2207189; Fax: 0039 041731330; E-mail: [sara.nordio@ospedalesancamillo.net](mailto:sara.nordio@ospedalesancamillo.net)

**Received** January 24, 2018; **Accepted** February 13, 2018; **Published** February 20, 2018

**Citation:** De Biagi F, Frigo AC, Andrea T, Sara N, Berta G, et al. (2018) Italian Validation of a Test to Assess Dysarthria in Neurologic Patients: A Cross-Sectional Pilot Study. *Otolaryngol (Sunnyvale)* 8: 343. doi: [10.4172/2161-119X.1000343](https://doi.org/10.4172/2161-119X.1000343)

**Copyright:** © 2018 De Biagi F, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

- “Protocollo di Valutazione della Disartria” [1]. In this regard, the following research questions are tested:

Are intra- and inter-scorer reliabilities for the protocol adequate for clinical purposes?

## Materials and Methods

### Study design

We used an experimental, cross-sectional pilot study for the validation of a tool to assess dysarthria in patients with neurological diseases. The Ethic Committee of Venice on 31<sup>st</sup> May 2016, reference number 49A/CESC, approved the study

### Subjects

50 dysarthric patients (28 males; 56%) participated in the study. All patients hospitalized at IRCCS San Camillo Hospital Foundation (Venice, Italy), diagnosed with dysarthria because of neurological etiology (August 2015-May 2016) and referred by the ward doctor for assessment of speech impairments were included.

The following exclusion criteria were considered:

- Inability to complete the protocol (e.g. severe cognitive disorders, aphasia, bucco-facial apraxia);
- Open tracheotomy tube;
- Italian not first language;
- Did not consent to examination being recorded via video.

After enrolment, patients were divided into two groups, according to their diagnosis. The first (n=25) was composed of patients diagnosed with degenerative dysarthria and the other with non-degenerative dysarthria (n=25). From August 2015 to May 2016 62 dysarthric patients were admitted to Hospital San Camillo and assessed by the speech and language therapy service. Six were excluded because the protocol was not feasible (among them 1 patient was diagnosed with aphasia, 3 patients suffered from bucco-facial apraxia and 2 other patients had severe cognitive disorders), 6 of them did not give consent to video recording. Table 1 summarizes descriptive statistic results of the 2 groups. Table 2 illustrates data of the included subjects [14-18].

### Scorers

Scoring was conducted by thirteen members of the speech and language therapy team at Hospital San Camillo, who were divided into two groups according to their work experience; “skilled” group comprised five speech and language therapists who have worked with dysarthria more than five years; “beginners” group consisted of eight speech and language therapists who had less than 5 years’ experience. Each participant was either on-line assessor or off-line scorer.

	Degenerative (n=25)	Non-degenerative (n=25)
Mean Age	58.4 ± 9.9	63.3 ± 11.1
Mean Time Post Onset (T.P.O.)	145.3 ± 126.1	62.4 ± 152.9
Diagnosis	7 MS, 7 PD, 5 ALS, 3 ataxia, 3 other	3 left stroke, 10 right stroke, 2 Arnold-Chiari Syndrome, 3 TBI, 2 Subarachnoid haemorrhage, 5 other

Table 1: Descriptive statistics of the two groups of patients.

N	Sex	Age (Years)	Diagnosis	Group	T.P.O. (Months)	Dominance
1	F	40	Multiple Sclerosis	1	72	R
2	M	81	Left Stroke	2	0.5	L
3	M	49	Cerebellar Ataxia	1	216	L
4	M	44	Arnold-Chiari Syndrome	2	300	L
5	M	51	Amyotrophic Lateral Sclerosis	1	72	L
6	F	60	Amyotrophic Lateral Sclerosis	1	180	L
7	M	79	Parkinson's Disease	1	108	L
8	M	72	Multiple Sclerosis	1	348	R
9	F	62	Amyotrophic Lateral Sclerosis	1	48	L
10	M	60	Multiple Sclerosis	1	84	L
11	M	61	Subarachnoid Hemorrhage	2	6	R
12	F	52	Multiple Sclerosis	1	408	R
13	F	50	Multiple Sclerosis	1	408	R
14	M	56	Left Stroke	2	1.5	L
15	F	58	Multiple Sclerosis	1	240	L
16	M	56	Traumatic Brain Injury	2	1.5	R
17	M	49	Parkinson's Disease	1	60	L
18	M	65	Parkinson's Disease	1	48	L
19	M	54	Amyotrophic Lateral Sclerosis	1	22	L
20	F	41	Multiple Sclerosis	1	312	L
21	M	73	Left Stroke	2	152	L
22	F	60	Parkinson's Disease	1	36	L
23	F	61	Multiple System Atrophy	1	60	L
24	F	71	Parkinson's Disease	1	36	L
25	M	45	Parkinson's Disease	1	96	L
26	F	70	Parkinson's Disease	1	59	L
27	F	74	Cerebellar Ataxia	1	36	L
28	M	61	Parkinson's Disease	1	120	R
29	F	73	Subarachnoid Hemorrhage	2	100	L
30	M	55	Bilateral Stroke	2	12	R
31	F	54	Amyotrophic Lateral Sclerosis	1	120	R
32	F	66	Cerebellar Ataxia	1	384	L
33	F	62	Arnold-Chiari Syndrome	2	147	L
34	M	62	Right Stroke	2	0.5	R
35	M	58	Progressive Supranuclear Palsy	1	60	L
36	F	37	Meningo-Cerebellitis	2	12	R
37	M	52	Right Stroke	2	3	L
38	F	63	Right Stroke	2	0.5	L
39	M	59	Traumatic Brain Injury	2	6	L
40	F	75	Right Stroke	2	4	L
41	M	53	Traumatic Brain Injury	2	11	L
42	M	70	Right Stroke	2	26	L
43	F	61	Cerebral Palsy	2	732	L
44	M	79	Right Stroke	2	3	L
45	M	73	Right Stroke	2	8	R
46	M	74	Right Stroke	2	2	L
47	F	51	Left Stroke	2	23	L
48	M	66	Right Stroke	2	1.5	L
49	M	79	Right Stroke	2	3	L
50	F	69	Guillan-Barré Syndrome	2	4	L

Dominance: L=Left; R=Right; Sex: M=Male; F=Female

Table 2: Data of included subjects.

## Procedure

The study consisted of 4 phases:

- **Phase 1: FOCUS GROUP AND TRAINING** (June-July 2015): the protocol was modified by means of focus group by the Speech and Language Therapy team following a literature review. Once the protocol was completed, the main researcher (DBF) administered the first assessment, which was video recorded. The groups were trained on the assessment and scoring methods by analyzing the video.
- **Phase 2: ENROLMENT** (August 2015-May 2016): 50 participants were recruited on the basis of the inclusion/exclusion criteria; assessments were video recorded. During this phase, data were stored in an anonymized form; an alphanumeric code was attributed to each participant.
- **Phase 3: PROTOCOL VALIDATION** (March-July 2016): Each video recording was scored by independent scorers, twice by the first author (DBF) to determine intra-rater reliability and once by another Speech and Language Therapist to determine inter-rater reliability.
- **Phase 4: STATISTICAL ANALYSIS** (July 2016): Intra-rater and inter-rater agreement were evaluated by means of the Lin's coefficient [20-24], with 95% confidence intervals. A lower limit of the 95% confidence greater or equal to 0.8 was considered indicating a good agreement. Analysis was performed with SAS 9.4 (SAS Institute, Inc., Cary, NC, USA).

## Materials

The protocol (attached 1) is a modified short-form of "Profilo di Valutazione della Disartria" [1] and is aimed to a perceptual analysis of the components that allow speech production. It is divided into seven subscales, each composed of a different number of items: intelligibility, respiration, phonation, diadochokinesis, oral muscles, prosody and articulation. Differently from the original protocol, one subscale ("reflexes") was completely removed and the total number of items was reduced from 71 to 35. This was done according to an internal construct validity study done by authors of the original scale, in order to shorten administration time. The score system of the original protocol was maintained in order to use the same normative data, that follow a 4-point Lickert scale (1=severe; 2=moderate; 3=mild; 4=normal) for each of the subscales described below.

**Subscale A: Intelligibility (2 items):** Contextual intelligibility was assessed through a brief sample of spontaneous speech; signal-dependent intelligibility is evaluated through a brief excerpt of reading.

**Subscale B: Respiration (3 items):** Two items were used to evaluate expiratory (prolonged /s/) and phonatory (prolonged /a/) durations. One item assessed the degree of pneumonic-phonatory coordination.

**Subscale C: Phonation (1 item):** Patients were asked to self-assess the degree of fatigue while speaking following 4 points Lickert scale. The assessor also provided qualitative data concerning voice production (intensity, voice quality).

**Subscale D: Diadochokinesis (6 items):** Patients were asked to repeat rapidly and accurately six different syllables. Scores included the number of syllables/5 s).

**Subscale E: Oral muscles (16 items):** Muscular functionality of lips, tongue, jaw and soft palate were assessed in terms of motility, range of

movements, rate and precision. Muscular strength was not included in the statistical analysis, because it was not possible to measure strength from a video recording.

**Subscale F: Prosody (4 items):** Two items assessed rhythm: patients were asked to repeat an automatic series (mounts of the year) at a normal and a faster rate. Two items assessed prosody: one item assessed the use of a normal intonation while speaking; another item assesses the ability of the patient to imitate different accents.

**Subscale G: Articulation (3 items):** Two items assessed the articulation and co-articulation of initial consonants and groups of consonants in the repetition of 44 words. One item assessed repetition of whole words (6 stimuli).

## Results

### Administration

All the scorers felt confident with both the administration and the scoring method of the protocol and no re-training was necessary. Administration time for the protocol was quite quick, ranging from a minimum of 8.4 min to a maximum of 30.1 min, with an average time of 17.1 min (SD=4.1). Almost all the participants succeeded to complete the protocol in only one session, but one patient had to schedule a second appointment due to fatigue. All participants, regardless of the severity and characteristics of the dysarthria, were able to perform almost all the subtests and items of the protocol.

### Video-recording scoring: Intra-rater and inter-rater reliability

The video-recording intra-rater agreement was evaluated comparing two protocol scorings (A1 and A2) for each subject done by the first author at two different times (t1 and t2). In order to avoid familiarity with the assessments and the patients, the two assessments were done at a distance of 1 month.

The inter-rater agreement was estimated by analyzing the scores (A1, A3 and A4) from three different scorers, for each patient. A1 was the score attributed from the first author; A3 was provided by one of the "expert" scorers and A4 was done by one of the "beginners".

The Concordance Correlation Coefficient (CCC) was estimated by means of the Lin's coefficient with 95% Confidence Intervals (95% CI). Analyses were performed taking into account the total scores of the seven subscales of the protocol. Moreover items 1 and 2 (respectively "Contextual intelligibility" and "Signal-dependent intelligibility") were also considered independently due to their clinical specificity as functional outcome measure for dysarthria severity.

The video-recording intra-rater and inter-rater agreement results are reported in Table 3.

For each subscale the intra-rater agreement was satisfactory (Figure 1). All the subscales had a CCC higher than 0.8 with a narrow CI. Almost all the CI upper limits were above 0.9 and for only one subscale ("Prosody") the lower limit was lower than 0.7. Subscale C "Phonation" had a perfect concordance because it was a self-reported measure; however this measure seemed unstable when assessed by different scorers.

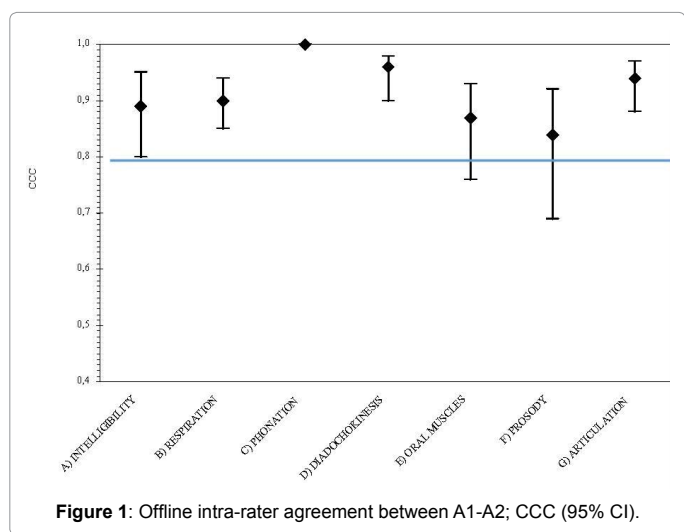
There was a high inter-rater agreement only for three of the subscales: B "Respiration", C "Phonation" and D "Diadochokinesis" (Figure 2). For each subscale the CI was wider than for the intra-rater agreement and for one subscale (A "Intelligibility") the CI lower limit was even lower than 0.5.

SUBSCALE		Offline Intra-rater agreement (A <sub>1</sub> -A <sub>2</sub> )		Offline Inter-rater agreement (A <sub>1</sub> -A <sub>3</sub> -A <sub>4</sub> )	
		N	CCC (95%CI)	N	CCC (95%CI)
A	Intelligibility	49	0.89 (0.80-0.95)	49	0.63 (0.48-0.74)
	- Signal-dependent	49	0.85 (0.71-0.94)	49	0.59 (0.43-0.72)
	- Contextual	50	0.81 (0.70-0.92)	50	0.57 (0.43-0.67)
B	Respiration	49	0.90 (0.85-0.94)	47	0.88 (0.83-0.93)
C	Phonation	45	1	42	0.89 (0.76-0.96)
D	Diadochokinesis	46	0.96 (0.90-0.98)	45	0.81 (0.73-0.87)
E	Oral muscles	46	0.87 (0.76-0.93)	42	0.75 (0.60-0.87)
F	Prosody	49	0.84 (0.69-0.92)	48	0.72 (0.63-0.86)
G	Articulation	49	0.94 (0.88-0.97)	49	0.74 (0.63-0.84)

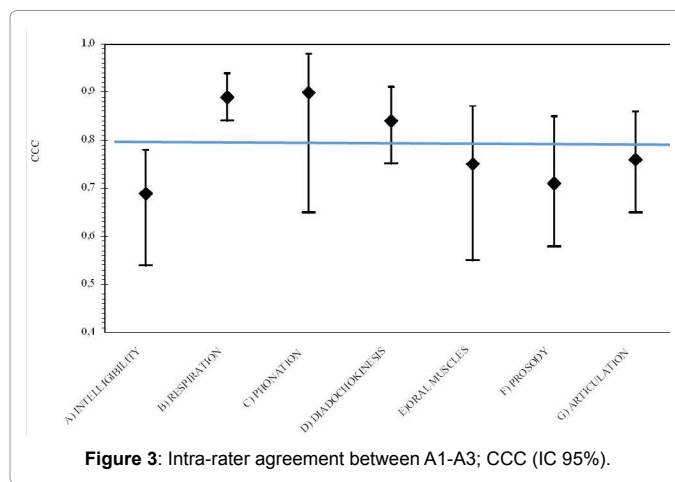
**Table 3:** Video-recording intra-rater and inter-rater agreement; concordance correlation coefficient (CCC) and 95% confidence intervals (95% CI).

Subscale		Offline Inter-rater agreement (A <sub>1</sub> -A <sub>3</sub> )		Offline Inter-rater agreement (A <sub>1</sub> -A <sub>4</sub> )	
		N	CCC (CI 95%)	N	CCC (CI 95%)
A	Intelligibility	49	0.69 (0.54-0.78)	49	0.67 (0.50-0.80)
	- Signal-dependent	49	0.68 (0.53-0.80)	49	0.57 (0.34-0.74)
	- Contextual	50	0.61 (0.48-0.74)	50	0.65 (0.48-0.79)
B	Respiration	47	0.89 (0.84-0.94)	48	0.87 (0.80-0.92)
C	Phonation	43	0.90 (0.65-0.98)	43	0.87 (0.73-0.94)
D	Diadochokinesis	45	0.84 (0.75-0.91)	46	0.73 (0.61-0.83)
E	Oral muscles	43	0.75 (0.55-0.87)	44	0.77 (0.60-0.89)
F	Prosody	48	0.71 (0.58-0.85)	49	0.73 (0.60-0.84)
G	Articulation	49	0.76 (0.65-0.86)	49	0.67 (0.52-0.80)

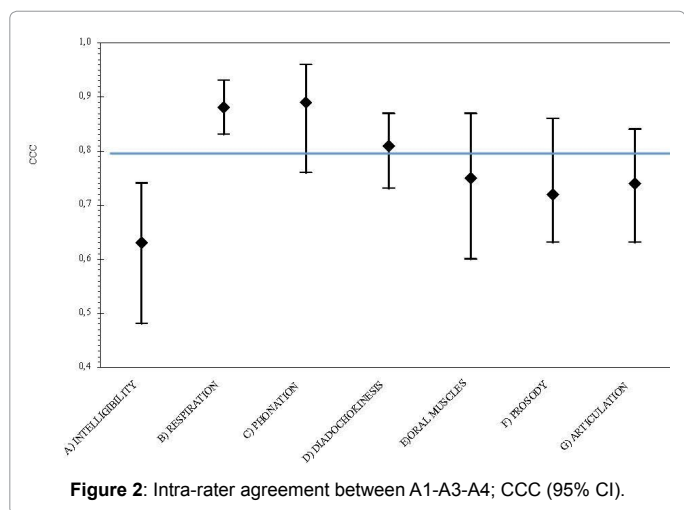
**Table 4:** Offline inter-rater agreement, difference between “skilled” and “beginner” scorers; concordance correlation coefficient (CCC) and 95% confidence intervals (95% CI).



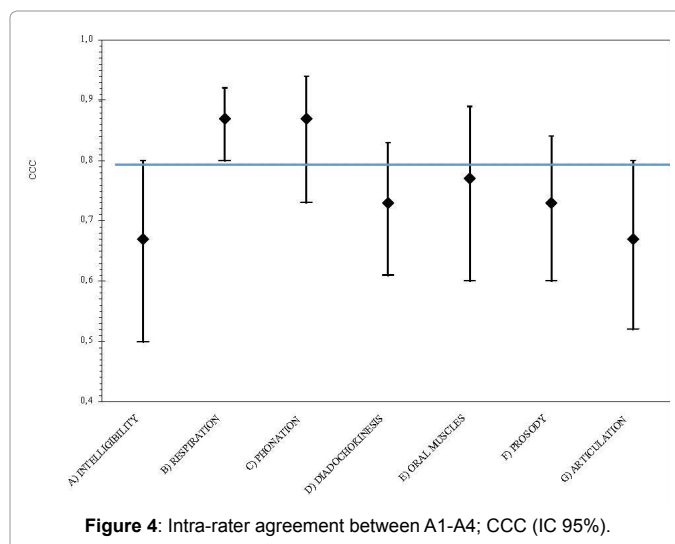
**Figure 1:** Offline intra-rater agreement between A1-A2; CCC (95% CI).



**Figure 3:** Intra-rater agreement between A1-A3; CCC (IC 95%).



**Figure 2:** Intra-rater agreement between A1-A3-A4; CCC (95% CI).



**Figure 4:** Intra-rater agreement between A1-A4; CCC (IC 95%).

The inter-rater agreement was also analyzed in relation to the expertise and acquaintance with the protocol. The CCC was estimated between the scores given by the first author (A1) and a “Skilled” SLT (A3) and between a “Beginner” SLT (A4). These analyses were performed in order to understand if the knowledge of the protocol could have affected the ability to rate. Table 4 displays the results. As it is shown in Table 4, there was a high CCC between the first author and “beginner” SLTs in only 2 subscales of the protocol (“Respiration” and “Phonation”). Whereas between a “skilled”

SLT and the first author, the agreement was satisfactory for three subscales (“Respiration”, “Phonation” and “Diadochokinesis”). Figures 3 and 4 show that the level of agreement was higher than 0.5 for all the items and for “Respiration” it was better than 0.8.

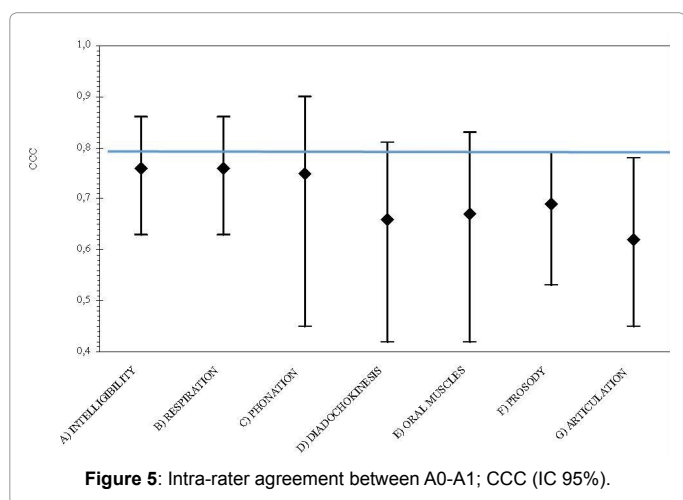
#### Face-to-face scoring: Inter-rater reliability

The intra- and inter-rater reliability was assessed analyzing the data



Online/offline inter-rater agreement ( $A_0-A_1$ )		
Subscale	N	CCC (CI 95%)
Intelligibility	49	0.76 (0.63-0.86)
- Signal-dependent	49	0.69 (0.59-0.81)
- Contextual	50	0.73 (0.51-0.86)
Respiration	49	0.76 (0.61-0.87)
Phonation	43	0.75 (0.45-0.90)
Diadochokinesis	46	0.66 (0.42-0.81)
Oral muscles	46	0.67 (0.42-0.83)
Prosody	49	0.69 (0.53-0.79)
Articulation	49	0.62 (0.45-0.78)

**Table 5:** Online/offline inter-rater agreement; concordance correlation coefficient (CCC) and 95% confidence interval (CI).



**Figure 5:** Intra-rater agreement between A0-A1; CCC (IC 95%).

from offline scoring, which were obtained by watching the patients in a video recording. Although the protocol is a face-to-face assessment the procedure based on video recording was designed with the aim to replicate a feasible setting within the SLT. Thus, we analyzed the online inter-rater reliability by estimating the CCC between online assessment (A0) and one of the offline scorings (A1). A1 was chosen because it the assessment was done by the first author as more experienced in the assessment and scoring. Table 5 exhibits the results.

As shown in Figure 5, all of the subscales of the protocol showed a poor agreement between the two scoring modalities (face-to-face/video-recording). The CI was wider and, although the entire CI upper limits were above 0.7, the lower limits could be even below 0.5.

## Discussion

### Psychometric features

The study aimed to explore the reliability of the protocol, analyzing the consistency in measurements between the same rater and among different raters with different level of expertise.

The intra-rater concordance was found to be very high, with a CCC more than 0.8 for many subscales and a tight CI. One subscale (“Phonation”), which was the only self-reported measure, even performed a perfect concordance. Qualitatively there were no significant differences among the performances of the different subscales.

Although the results of the intra-rater reliability should always be interpreted with respect to the skill-level of the clinician [19], these data that were not calculated in the original protocol [13] showed

that the scoring system is stable in measurements repeated over time. Thus, the results suggested that the protocol could be a reliable tool to track patient’s progress. According to the literature [20], this point is crucial given the importance of an objective measure to determine the effectiveness of the treatment.

However, the inter-rater agreement was found to be worse than the intra-rater one. In fact, only three subscales of the protocol had a high CCC. Not surprisingly, two of these subscales (“Respiration” and “Diadochokinesis”) are the most objective measurements, with clear normative data for scoring. Further, one of these three subscales is a self-reported measure. Nevertheless, from a qualitative perspective, the agreement for the other subscales could be considered satisfactory as well. Generally, the CCC for the all other subscale ranges from 0.63 to 0.75, with almost all the CI lower limits above 0.6 and all the upper limits all above 0.8. One exception is “Intelligibility”, which was the subscale that had the worst agreement, with a CI lower limit less than 0.5. This result could be discouraging considering the significance of this parameter that should be considered as the main functional outcome of speech disturbances rehabilitation. Indeed, previous literature highlighted that the assessment of intelligibility is a well-established problem in the field [21] and the dispute on how to measure is no nearer to be solved [10].

Our data indicate that four out of seven are susceptible to subjective judgment. Thus, it is possible that the scoring system or the measurement of some items should be revised in order to make it as more objective as possible.

Other analyses were performed in order to understand whether the discrepancies between inter and intra-rater reliability could have been related to the expertise of the raters. However, no significant difference was found comparing expert to non-skilled raters. Thus inconsistency can be attributed to reasons other than lack of training or knowledge of the instrument by scorers.

Furthermore, we found a poor inter-rater agreement between online and offline scoring. This may raise questions about the suitability conducting this assessment via video recording rather than face-to-face. This finding is inconsistent with a part of literature that explores the strength of agreement between face-to-face and off-line evaluation of dysarthric speech: Hill et al. [22] founded a high intra-rater and inter-rater reliability within both the assessment methods.

Nevertheless, future studies are required to evaluate the effectiveness of assessment of speech based on face-to-face or off-line methods.

### Clinical utility

The results supported a good clinical utility of the protocol.

Administration time for the protocol is short, which suggest that it is suitable for clinical practice. Moreover, administration time was also sufficient enough in relation to patients’ abilities. In fact, almost all the included subjects managed to complete the assessment in only one session. These results are satisfactory and confirmed the need to use a short-form of the original protocol. Moreover these results are consistent with the findings provided by the rush analysis performed by the authors of the protocol [13].

The protocol does not require any technical or specific equipment. The only resource that should be accounted for the administration is the time of the healthcare professional, which is strongly recommended to be a trained SLT.

In addition, the protocol seemed easy to administer; none of our scorers needed to be retrained. Our training was achieved in a couple of sessions lasting a few hours. These two features also imply limited organizational constraints.

Finally, the difficulty of the protocol's items seemed to be adequate for both the different kinds of dysarthria and the various severity levels. Almost all the included subjects were able to complete almost all the items.

## Limitations

The aim of this study was to conduct an exploratory pilot study. Thus, the generalizability of our findings is limited.

Increasing sample size and diversity in the types of clients examined, together with the recruitment of raters from other centers, might be useful for minimizing the possible bias in future studies.

Moreover it should be taken into consideration the limitations of the clinical utility of the protocol as a clinician-related measure. In fact, the client's perspective and expectations are not considered. Moreover, the protocol is aimed to assess speech impairment, while the levels of disability and participation are not considered. This may be in contrast with the latest research that emphasizes participation-focused assessments and interventions.

## Conclusion

Even though further analyses are required, our preliminary data showed a good consistency of ratings in repeated measures over time. Nevertheless, ratings between different scorers were less stable, especially in the face-to-face administration of the protocol and many items of the protocol were found to be susceptible to individual judgment. Above all, the measurement of the functional outcome (i.e., intelligibility) seemed to be unsatisfactory. The discrepancy between inter- and intrarater reliability was not attributed to the level of acquaintance with the protocol. Thus, it may be postulated that the scoring system itself, as well as the normative data, should be reconsidered.

The study revealed an adequate clinical utility of the protocol, whose administration has been thought to be convenient and affordable in terms of duration and resources required. Besides, the protocol with regard to parameters assessed and items' difficulty seemed to be adequate for the different type and level of dysarthria severity.

In conclusion, while the protocol appears to be a potentially useful test, the study warrants cautious interpretation, due to the limited generalizations of the findings. Further research is indeed required in order to validate the instrument, possibly integrating it with other types of outcome measures.

Future studies are needed to foster the use of standardized and validated tool to assess outcomes in rehabilitation. A grounded measurement of the outcomes is undeniably important in order to establish the patient's baseline status and monitor his improvements, determining the usefulness of treatments.

In this way, outcome measures reporting may contribute to improve clinical practice, supporting organizational changes and leading to efficient policy acts.

## Acknowledgment

This project benefited from the help, support and guidance of many persons.

In addition to the mentioned authors of the article, a heartfelt thanks to the SLT team at Hospital San Camillo: Giulia Berta, Sara Nordio, Virginia Reay, Sara Ranaldi, Martina Garzon, Irene Battel, Isabella Koch, Jessica Fundarò, Silvia Biasin, Federica Biddau and Alessandra Ceolin.

## Declaration of Interests

The authors declare that there is no conflict of interests.

## References

1. Fussi F, Cantagallo A (1997) Profilo di valutazione della disartria. Torino: Omega.
2. Duffy JR (2013) Defining, understanding and categorizing motor speech disorders. In: Duffy JR (eds.) motor speech disorders. substrates, differential diagnosis and management. St. Louis: Motsby.
3. Wang EQ (2010) Dysarthria. In: Kompolti K, Metman LV (eds.) Encyclopedia of movement disorders. San Diego, CA: Elsevier LTD.
4. Baylor C, Burns M, Aedie T, Britton D, Yorkston K (2011) A qualitative study of interference with communicative participation across communication disorders in adults. *Am J Speech Lang Pathol* 20: 269-287.
5. Mackenzie C, Muir M, Allen C, Jensen A (2014) Non-speech oro-motor exercises in post-stroke dysarthria intervention: A randomized feasibility trial. *Int J Lang Commun Disord* 49: 602-617.
6. Collins J, Bloch S (2012) Survey of speech and language therapists' assessment and treatment practices for people with progressive dysarthria. *Int J Lang Commun Disord* 47: 725-737.
7. Convey A, Walshe M (2015) Management of non-progressive dysarthria: Practice patterns of speech and language therapists in the Republic of Ireland. *Int J Lang Commun Disord* 50: 374-388.
8. Haynes WO, Pinzola RH (2011) Diagnosis and evaluation in speech pathology.
9. Scettino I, Albino F, Ruoppolo G (2013) La valutazione bedside del soggetto disartrico. In: Ruoppolo G, Amitrano A (eds.) A cura di. Disartria: Possiamo fare di più? Relazione 2013 Ufficiale Congresso nazionale della Società Italiana di Foniatria e Logopedia. Torino: Omega Edizioni, pp: 31-42.
10. Miller N (2013) Measuring up to speech intelligibility. *Int J Lang Commun Disord* 48: 601-612.
11. Enderby P (1980) Frenchay dysarthria assessment. *Br J Disord Commun* 21: 165-173.
12. Robertson SJ (1982) Dysarthria profile. Tucson: Communication Skill Builders.
13. Cantagallo A, La Porta F, Abenante L (2006) Dysarthria assessment: Robertson profile and self-assessment questionnaire. *Acta Phoniatria Latina* 28: 246-261.
14. Lin L, Hedayat AS, Sinha B, Yang M (2002) Statistical methods in assessing agreement. *J Am Statist Assoc* 97: 257-270.
15. Lin L, Hedayat AS, Wu W (2012) Statistical tools for measuring agreement. New York: Springer.
16. Lin L, I-kuei L (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45: 255-268.
17. Lin L, I-kuei L (1992) Assay validation using the concordance correlation coefficient. *Biometrics* 48: 599-604.
18. Lin L, I-kuei L (2002) A note on the concordance correlation coefficient. *Biometrics* 56: 324-325.
19. Bialocerkowsky AE, Bragge P (2008) Measurement error and reliability testing: Application to rehabilitation. *Int J Ther Rehabil* 15: 422-427.
20. Yorkston KM (1996) Treatment efficacy: Dysarthria. *J Speech Hear Res* 39: S46-557.
21. Berisha V, Utianski R, Liss J (2013) Towards a clinical tool for automatic intelligibility assessment. *Proc IEEE Int Conf Acoust Speech Signal Process* 2013: 2825-2828.
22. Hill AJ, Theodoros DG, Russel TG, Ward EC (2009) The redesign and re-evaluation of an internet-based telerehabilitation system for the assessment of dysarthria in adults. *Telemed J E Health* 15: 840-850.