

HEAD AND NECK

Validity and reliability of the Italian translation of the Yale Pharyngeal Residue Severity Rating Scale

Studio di validazione e affidabilità della traduzione italiana della Yale Pharyngeal Residue Severity Rating Scale

Sara Nordio¹, Lorenza Maistrello¹, Daniela D'Imperio¹, Niccolò Favaretto², Angela Dellai³, Silvia Montino², Anna Agostinelli², Giulia Ramacciotti², Flavia Gheller⁴, Giulia Berta¹, Isabella Koch¹, Elisabetta Zanoletti², Irene Battel^{5,6}, Diego Cazzador²

¹IRCCS San Camillo Hospital, Venice, Italy; ²Department of Neuroscience, Section of Otolaryngology, Azienda Ospedale-Università Padova, Padua, Italy; ³Unità Operativa Medicina Fisica e Riabilitativa, Ulss 6 Euganea, Ospedale di Conselve, Padua, Italy; ⁴Department of Developmental Psychology and Socialisation (DPSS), University of Padua, Padua, Italy; ⁵Unità Operativa Medicina Fisica e Riabilitativa, Ulss 3 Serenissima, Ospedale Civile di Venezia, Venice, Italy; ⁶Department of Clinical Speech & Language Studies, University of Dublin, Trinity College, College Green, Dublin 2, Ireland

SUMMARY

Objective. In the dysphagic patient, pharyngeal residues (PR) are associated with aspiration and poor quality of life. The assessment of PR using validated scales during flexible endoscopic evaluation of swallowing (FEES) is crucial for rehabilitation. This study aims to validate and test the reliability of the Italian version of the Yale Pharyngeal Residue Severity Rating Scale (IT-YPRSRS). The effects of training and experience in FEES on the scale were also determined.

Methods. The original YPRSRS was translated into Italian according to standardised guidelines. Thirty FEES images were selected after consensus and proposed to 22 naive raters who were asked to assess the severity of PR in each image. Raters were divided into two subgroups by years of experience at FEES, and randomly by training. Construct validity, inter-rater, and intra-rater reliability were assessed by kappa statistics.

Results. IT-YPRSRS showed substantial to almost perfect agreement ($\kappa > 0.75$) in validity and reliability for both the overall sample (660 ratings), and valleculae/pyriform sinus sites (330 ratings each). No significant differences emerged between groups considering years of experience, and variable differences were observed by training.

Conclusions. The IT-YPRSRS demonstrated excellent validity and reliability in identifying location and severity of PR.

KEY WORDS: deglutition, fiberoptic endoscopic evaluation of swallowing, pharyngeal residue; reliability, scale

RIASSUNTO

Obiettivo. Nel paziente disfagico i residui faringei (RF) correlano con rischio di aspirazione e bassa qualità di vita. La loro valutazione durante FEES attraverso scale validate è fondamentale. Questo studio mira a validare e testare l'affidabilità della versione italiana della Yale Pharyngeal Residue Severity Rating Scale (IT-YPRSRS), una scala che valuta i RF. Vengono inoltre studiati gli effetti di esperienza professionale e training.

Metodi. La YPRSRS in versione inglese è stata tradotta in Italiano secondo protocollo standardizzato. Trenta immagini post-deglutitorie sono state selezionate da tre esperti nell'uso della YPRSRS e proposte a 22 valutatori naif, che hanno determinato il livello di gravità dei RF in ciascuna. I valutatori sono stati quindi divisi in due sottogruppi secondo: anni di esperienza nella FEES e training. La validità di costruito e l'affidabilità intra- ed inter-operatore sono state studiate con kappa statistics.

Risultati. La IT-YPRSRS ha mostrato un elevato grado di agreement ($\kappa > 0,75$) in termini di validità di costruito e di affidabilità, per la totalità del campione (660 ratings) e per le sedi di vallecole e seni piriformi separatamente (330 ratings ciascuno). Non sono state rilevate differenze significative per gli anni di esperienza, mentre si è osservato un miglior agreement nei risultati del gruppo sottoposto a training.

Received: August 30, 2022
Accepted: December 18, 2022

Correspondence

Diego Cazzador
Department of Neuroscience, Section of Otolaryngology, University of Padova, via Giustiniani 2, 35128 Padua, Italy
Tel. +39 049 821 8778. Fax +39 049 821 1994
E-mail: diego.cazzador@unipd.it

How to cite this article: Nordio S, Maistrello L, D'Imperio L, et al. Validity and reliability of the Italian translation of the Yale Pharyngeal Residue Severity Rating Scale. Acta Otorhinolaryngol Ital 2023;43:12-19. <https://doi.org/10.14639/0392-100X-N2289>

© Società Italiana di Otorinolaringoiatria e Chirurgia Cervico-Facciale



OPEN ACCESS

This is an open access article distributed in accordance with the CC-BY-NC-ND (Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International) license. The article can be used by giving appropriate credit and mentioning the license, but only for non-commercial purposes and only in the original version. For further information: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>

Conclusioni. La IT-YPRSRS ha dimostrato un eccellente livello di validità e affidabilità nell'identificazione e nella quantificazione di RF.

PAROLE CHIAVE: *deglutizione, FEES, residui faringei, affidabilità, scala*

Introduction

Pharyngeal residues (PR), defined as the retention of liquids or food in the pharynx after swallowing, is one of the most relevant signs of oropharyngeal dysphagia¹. This condition severely affects patients' morbidity, due to the high risk of aspiration, malnutrition and dehydration², thus decreasing patients' quality of life and social participation³. Accordingly, PR may be a predictor of post-swallow penetration and aspiration^{4,6}, so that the accurate identification and quantification of the residue severity in the clinical setting is crucial⁷.

Residues can be directly observed through fiberoptic endoscopic evaluation of swallowing (FEES), which allows identification of exact anatomical sites and amount⁴. FEES and videofluoroscopic swallow study (VFSS) are well-recognised methods for diagnosis and management of dysphagia⁸. Although FEES is more sensitive than VFSS in quantifying PR, this procedure can be affected by subjective interpretation of the clinician and compared to VFSS lacks objective tools for PR assessment⁹⁻¹¹. In addition, the simple ascertainment of the presence of PR can be insufficient for clinical and rehabilitation management, as well as for research purposes. To overcome these issues, several rating scales have been proposed to define dysphagia and PR severity during FEES^{4,12,13}. However, a systematic review by Neubauer et al.¹⁴ on PR severity rating scales based on FEES, revealed many methodological flaws in the included studies, and only the Yale Pharyngeal Residue Severity Rating Scale (YPRSRS)¹⁵ resulted a valid and reliable tool for residue assessment.

The YPRSRS is an anatomically defined, 5-points ordinal scale (none, trace, mild, moderate, severe) that outlines the residue severity patterns in the valleculae and pyriform sinuses¹⁵. This scale has already been validated in German¹⁶ and Turkish¹⁷. To the best of our knowledge, no such tools are available in Italian to date.

The importance of adopting the same internationally validated instruments for PR evaluation can facilitate the comparability of results across different countries. The cross-cultural translation process ensures that a translated measurement tool is understood in a cultural context that is different from the original setting and that does not lose its measurement properties^{18,19}.

This study aimed to: (i) translate the YPRSRS, (ii) assess the psychometric properties of the Italian version of the IT-YPRSRS, and (iii) determine if training and experience have an impact on the results of the IT-YPRSRS.

Materials and methods

A cross sectional design was selected for the present study. Research methodology recommends following a standardised translation and validation process of a scale, in order to achieve appropriate linguistic accuracy²⁰.

Translation

After authorisation by the authors of the YPRSRS (Neubauer PD, personal communication), the translation process included the following steps: (i) forward translation and its review for consensus; (ii) backward translations and its review for consensus. Three forward translations into Italian language were produced by two bilingual Otolaryngologists and two Speech and Language Pathologists (SLP) involved in the management of patients with dysphagia. The three versions were discussed and merged after consensus. Next, two external native English speakers with excellent Italian language skills performed two back translations of the consensus version into English. The back translations were subsequently compared to the original version and discussed by the expert committee (the Otolaryngologists and the SLPs) which stated all items of the scale were completely clear. The Italian version of the scale is reported in Appendix I.

Fiberoptic endoscopic evaluation of swallowing

FEES examination was conducted by otorhinolaryngologists together with SLPs using a flexible transnasal laryngoscope and Tele Pack system (KARL STORZ SE & Co. KG, Tuttlingen, Germany). Each examination was anonymously recorded as .AVI files. Fifty-five consecutive patients with swallowing impairments from neurogenic aetiology were recruited, with a prevalent diagnosis of stroke-induced dysphagia (56.4%), whereas the other diagnoses included brain injury, brain tumour, multiple sclerosis and amyotrophic lateral sclerosis. FEES was completed with the following bolus types, twice each: cracker for solid food (IDDSI 7), 5-mL yogurt for pureed food (IDDSI 4), and 5-mL milk for liquid food (IDDSI 2)²¹.

Image selection

Initially, 103 post-swallow images were selected from the recorded videos. All the frames were captured at the end of the first swallow to have homogeneous data for rating¹⁵. Ninety of these displayed bolus residues, while 17 images displayed no residues. The first step of the selection pro-

cess consisted in the categorisation of the 103 images for reference in severity rating. Given the absence of a gold standard reference, three FEES experts with a combined 23 years of expertise (range 6-9 years) independently conducted the task according to the YPRSRS severity. Only images with complete agreement among the three experts were included (64 images), thus obtaining one reference value per image for the calculation of the construct validity. Finally, a further image selection was conducted following the “best-of-the-best” criterion, and 30 images were chosen by consensus (15 for valleculeae and 15 for pyriform sinuses, with homogeneous distribution of scores, i.e. 3 images per each score class). Less-defined frames were excluded.

Rating process

Ten otorhinolaryngologists and 12 SLPs regularly involved in FEES, with a minimum professional experience of 3 years, were recruited at different Italian institutions. None of the 22 raters had ever used the English version of the YPRSRS for either clinical or scientific purposes. After agreeing to participate, raters were grouped according to training status and years of experience with FEES. As for training, raters were randomly assigned (by means of a computer-generated order using the appropriate Excel function) to receive or not a specific 4-minute training video in Italian. The video explained the rationale, application and clinical significance of the YPRSRS, and provided images (different from those selected for data collection) for each grade of severity. As for years of experience, raters were divided in two subgroups by the median value.

The colour images (15 for valleculeae evaluation, and 15 for pyriform sinuses) were sent via email as an editable pdf file, presenting one image per page at a resolution of 720 x 476 pixels. The file included the Italian version of the YPRSRS. Raters were asked to assess the severity level of residue for each image by selecting the considered appropriate value. A second round of rating was performed after 15 days to assess intra-rater reliability¹⁵. For this purpose, the same images were randomly rearranged in a new editable pdf file.

Statistics

Descriptive statistics were calculated to obtain the demographic and professional characteristics of the raters. Associations of discrete and continuous variables were assessed with Fisher’s exact test and Mann-Whitney U test, respectively. Construct validity, intra-rater reliability, and inter-rater reliability were calculated using kappa statistics, standard errors (SEs) and 95% confidence intervals (95% CIs)²². In particular, construct validity was determined with Cohen’s kappa coefficient with quadratic weights²³

on the agreement between the first evaluation of each rater and that of the experts. The intra-rater reliability was determined by calculating Cohen’s kappa coefficient with quadratic weights by the agreement between the first and the second rating. The degree of agreement across several raters (inter-rater reliability) was calculated by the Fleiss’ kappa with quadratic weights²⁴. The analyses were performed for the overall sample and for subgroups of raters to assess whether there was a difference in outcomes related to the level of experience, and training status.

To interpret the results, the criteria proposed by Landis and Koch were used for Cohen’s kappa: values between 0.41-0.60 represent moderate agreement, values between 0.61-0.80 substantial agreement, and values between 0.81-1.00 almost perfect agreement²³. The following benchmark was adopted for the Fleiss kappa: values < 0.40 poor agreement, between 0.40-0.75 intermediate to good, and > 0.75 excellent agreement²⁴. Kappa values of different subgroups were compared using Z-statistics¹⁵. All statistical analyses were performed using the IBM SPSS software for Windows, version 26.0 (SPSS Inc., Chicago, IL, USA). The significance level was set at 0.05.

Results

Characteristics of raters

Twenty-two raters, including 12 (55%) SLPs and 10 (45%) otorhinolaryngologists took part in the study. All raters completed the grading of FEES images in the given time, and none left the study. Participants’ characteristics are reported in Table I.

Measures of validity and reliability

Kappa statistic calculated on the entire sample of 660 ratings reported substantial to almost perfect agreement, with no values below 0.75 (Tab. II). Analyses performed on 330 ratings according to the anatomical location confirmed excellent degrees of validity and reliability (Tab. II). Specifically, construct validity was almost perfect for both anatomical sites (kappa = 0.98 ± 0.02). There was excellent inter-rater agreement for the valleculeae and pyriform sinuses locations, with kappa of 0.81 ± 0.01 and 0.78 ± 0.01 , respectively; and an almost perfect intra-rater agreement for both sites with kappa = 0.95 ± 0.01 in evaluation of valleculeae residues, and kappa = 0.93 ± 0.01 in assessment of pyriform sinus residues.

Ratings by years of experience

The 22 raters had a median value of 7 years of FEES experience. The 12 less-experienced raters had a median of 5 years experience at FEES (IQR 4.3-6.0), while the 10

Table I. Raters' characteristics.

	Total (n = 22)	Experience		P-value	Training status		P-value
		< 7 years (n = 12)	≥ 7 years (n = 10)		No (n = 11)	Yes (n = 11)	
Sex				1.000			0.090
Female, N (%)	18 (82%)	10 (56%)	8 (44%)		11 (61%)	7 (39%)	
Male, N (%)	4 (18%)	2 (50%)	2 (50%)		0 (0%)	4 (100%)	
Profession				0.691			1.000
SLP, N (%)	12 (55%)	6 (50%)	6 (50%)		6 (50%)	6 (50%)	
MD, N (%)	10 (45%)	6 (60%)	4 (40%)		5 (50%)	5 (50%)	
Years of experience, median [IQR]	7 [5.0-10.0]	5 [4.3-6.0]	12 [8.8-27.0]	-	6 [5.0-8.5]	8 [6.0-12.0]	0.373*

*Mann-Whitney U test; N: number of raters; SLP: speech and language pathologist; MD: otorhinolaryngologist; IQR: interquartile range.

Table II. Construct validity, inter-rater reliability, and intra-rater reliability kappa statistics for valleculeae and pyriform sinus ratings across all raters.

	Overall (N = 660)		Valleculeae (N = 330)		Pyriform sinus (N = 330)	
	Kappa (± SE)	95% CI	Kappa (± SE)	95% CI	Kappa (± SE)	95% CI
Construct validity	0.98 (± 0.01)	0.95; 1.00	0.98 (± 0.02)	0.95; 1.00	0.98 (± 0.02)	0.94; 1.00
Inter-rater reliability	0.79 (± 0.01)	0.79; 0.81	0.81 (± 0.01)	0.79; 0.83	0.78 (± 0.01)	0.77; 0.80
Intra-rater reliability	0.94 (± 0.01)	0.93; 0.95	0.95 (± 0.01)	0.94; 0.97	0.93 (± 0.01)	0.91; 0.95

N: number of ratings; SE: Standard Error; 95% CI: 95% confidence interval.

experienced raters had a median of 12 years (IQR 8.8-27.0) of experience (Tab. I).

Kappa statistic according to years of experience is reported in Table III. Construct validity, as well as intra-rater reliability showed almost perfect agreement for both anatomical locations. No significant differences were observed between groups. The inter-rater reliability for valleculeae anatomical site showed excellent agreement in both less-experienced and experienced raters (kappa = 0.79 ± 0.016 vs 0.83 ± 0.019, respectively; p = 0.184). For pyriform si-

nuses, inter-rater reliability agreement was excellent as well (kappa = 0.78 ± 0.016 vs 0.76 ± 0.019) without significant differences by experience at FEES.

Ratings by training status

Of the 22 rates involved in the study, 11 (50%) were randomly assigned to receive a specific training in the interpretation and application of the YPRSRS. There was no significant difference in years of FEES experience between trained vs non-trained groups (Tab. I).

Table III. Construct validity, inter-rater reliability, and intra-rater reliability kappa statistics for valleculeae and pyriform sinus ratings by years of experience.

	Experience < 7 years		Experience ≥ 7 years		P-value
	Kappa (± SE)	95% CI	Kappa (± SE)	95% CI	
Construct validity					
Valleculeae	0.98 (± 0.015)	0.95; 1.00	0.98 (± 0.015)	0.95; 1.00	0.805
Pyriform sinus	0.98 (± 0.017)	0.94; 1.00	0.97 (± 0.019)	0.93; 1.00	0.909
Inter-rater reliability					
Valleculeae	0.79 (± 0.016)	0.76; 0.83	0.83 (± 0.019)	0.79; 0.87	0.184
Pyriform sinus	0.78 (± 0.016)	0.75; 0.82	0.76 (± 0.019)	0.72; 0.79	0.314
Intra-rater reliability					
Valleculeae	0.96 (± 0.009)	0.95; 0.98	0.94 (± 0.012)	0.92; 0.97	0.142
Pyriform sinus	0.92 (± 0.014)	0.89; 0.95	0.93 (± 0.014)	0.90; 0.96	0.801

SE: Standard Error; 95% CI: 95% confidence interval.

Kappa statistic according to training status is reported in Table IV. Results showed almost perfect agreement for construct validity in both anatomical locations for both raters with and without training (valleculae, $p = 0.733$; pyriform sinus, $p = 0.892$).

Inter-rater reliability in valleculae assessment showed intermediate-to-good agreement ($\kappa = 0.72 \pm 0.018$) for the training group, and excellent agreement ($\kappa = 0.82 \pm 0.017$) for the subgroup that received training. Similarly, in pyriform sinus assessment, intermediate-to-good agreement was found for both the non-training group ($\kappa = 0.56 \pm 0.017$), and trained group ($\kappa = 0.68 \pm 0.018$). The differences in the degree of agreement by training status were significant for PR assessment in both anatomical sites ($p < 0.001$).

Intra-rater reliability showed almost perfect agreement by training status. In particular, considering valleculae residue assessment, trained raters demonstrated a significantly higher agreement than non-trained participants ($\kappa = 0.98 \pm 0.006$ vs 0.93 ± 0.013 , $p < 0.001$).

Discussion

The identification and quantification of PR severity during FEES is of utmost importance in the management of patients with swallowing disorders^{1,6,12}. Direct endoscopic documentation of PR is currently integrated by the use of validated scales for assessment of PR severity^{4,12-15}. First presented in 2015¹⁵, the YPRSRS is an anatomically defined tool for evaluation of residue location (valleculae and pyriform sinuses) and residue amount (5-point scale from none to severe). Among different scales proposed in the literature, the YPRSRS is probably the most valid and reliable, according to a recent systematic review that aimed to compare the qualitative and psychometric properties of

FEES-based PR scales¹⁴. Therefore, in this study we aimed to translate and test the psychometric properties of the Italian version of the YPRSRS.

The IT-YPRSRS showed excellent construct validity and intra-rater reliability, as well as high inter-rater reliability for both anatomical sites, valleculae and pyriform sinuses (Tab. II). These findings were consistent with that presented in the original validation of the English version of the scale¹⁵. The comparable results confirmed the appropriateness of the IT-YPRSRS and its potential inclusion in an Italian clinical context.

Similar to what reported for the original YPRSRS¹⁵ as well as in the validation of the German and Turkish versions^{16,17}, the current investigation observed that experience does not influence PR ratings. In particular, we calculated similar levels of agreement in both the experienced (median < 7 years) and non-experienced (median ≥ 7 years) groups of raters, for all psychometric properties analysed (Tab. III), suggesting that high competence in the use of the IT-YPRSRS can be achieved by SLPs and physicians with different levels of expertise.

In the IT-YPRSRS, significantly different values of kappa for the valleculae and pyriform sinuses were found for inter-rater reliability between trained and non-trained raters, with higher kappa registered in the former group ($p < 0.001$). Analogously, higher kappa intra-rater agreement was calculated for valleculae PR assessment by trained raters ($p < 0.001$). These results coincided with previous data from Neubauer¹⁵ and Gerschke¹⁶, confirming that rating precision might benefit from minimal training before the YPRSRS is applied to patients. We randomly proposed to raters a brief 4-minute training video that explained the use and rationale of the YPRSRS. This video can be adopted in clinics for practitioners who approach the IT-YPRSRS for the first time.

Table IV. Construct validity, inter-rater reliability, and intra-rater reliability kappa statistics for valleculae and pyriform sinus ratings by training received in the application of the Yale scale.

	No training received		Training received		p-value
	Kappa (\pm SE)	95% CI	Kappa (\pm SE)	95% CI	
Construct validity					
Valleculae	0.98 (\pm 0.012)	0.96; 1.00	0.98 (\pm 0.017)	0.94; 1.00	0.733
Pyriform sinus	0.98 (\pm 0.017)	0.94; 1.00	0.97 (\pm 0.019)	0.94; 1.00	0.892
Inter-rater reliability					
Valleculae	0.72 (\pm 0.018)	0.68; 0.75	0.82 (\pm 0.017)	0.79; 0.86	< 0.001
Pyriform sinus	0.56 (\pm 0.017)	0.53; 0.59	0.68 (\pm 0.018)	0.64; 0.71	< 0.001
Intra-rater reliability					
Valleculae	0.93 (\pm 0.013)	0.90; 0.95	0.98 (\pm 0.006)	0.97; 0.99	< 0.001
Pyriform sinus	0.91 (\pm 0.015)	0.88; 0.94	0.95 (\pm 0.012)	0.92; 0.97	0.054

SE: Standard Error; 95% CI: 95% confidence interval.

Finally, some limitations of this study need to be noted. First, as already mentioned in the German Yale validation study¹⁶, the pre-testing step of the validation process of the IT-YPRSRS was conducted by consensus with FEES experts²⁰. Second, as Neubauer and colleagues did in the original research¹⁵ we selected FEES frames for PR severity ratings. Undeniably, frames do not represent a real-life clinical setting, as FEES videos might be more appropriate. However, a recent study based on YPRSRS recognised the substantial complexity in rating PR on videos rather than on frames. The study demonstrated a trend of lower psychometric properties for videos, in comparison to an almost perfect agreement for frames²⁵. Third, the “best-of-the-best” criterion adopted for the selection of the final images pool denoted a limitation, as it is not representative of common clinical practice. This might have led to increased levels of agreement in our study. Fourth, three different consistencies were randomly used in the study, which on one hand represents only a part of boluses available for FEES²¹, and on the other this might have lowered agreement results. In fact, it has been recently demonstrated that bolus consistencies play a role in determining the psychometric properties of the YPRSRS, with thin liquids (IDDSI 0) having the lowest levels of rating agreement than solid food (IDDSI 7) and pureed food (IDDSI 4), respectively²⁵. Specific analysis for each consistency rated with the IT-YPRSRS and its clinical correlation is desirable in future research to better target rehabilitative interventions.

Conclusions

The psychometric characteristics of the IT-YPRSRS make it a validated, reliable and valuable tool to integrate FEES evaluation. This anatomically defined instrument is easy to administrate with a minimum training, regardless of years of experience of practitioners. We hope that the dissemination of the IT-YPRSRS will contribute to improve the accuracy of FEES, given the crucial role of characterising valleculae and pyriform sinus residues in the dysphagic patient.

Acknowledgments

The authors would like to thank Paul D. Neubauer, Yale School of Medicine, for his support.

Conflict of interest statement

The authors declare no conflict of interest.

Funding

This work was supported by the Italian Ministry of Health (Ricerca Corrente).

Author contributions

All authors contributed to the study conception and design. SN, NF, FG, GR, SM, DC, GB, IK, AD: material preparation and data collection; LM, DDI, DC: data analysis and results; SN: first draft of the manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethical considerations

The study was approved by the Ethics Committee of the hospital (Prot.2018.14-YALE) and carried out in accordance with the Declaration of Helsinki. All participants signed an informed consent for their inclusion.

References

- Logemann JA. Evaluation and treatment of swallowing disorders. Second Edition. Austin, Texas: PRO-ED; 1998.
- Marik PE. Pulmonary aspiration syndromes. *Curr Opin Pulm Med* 2011;17:148-154. <https://doi.org/10.1097/MCP.0b013e32834397d6>
- Jones E, Speyer R, Kertscher B, et al. Health-related quality of life and oropharyngeal dysphagia: a systematic review. *Dysphagia* 2018;33:141-172. <https://doi.org/10.1007/s00455-017-9844-9>
- Murray J, Langmore SE, Ginsberg S, et al. The significance of accumulated oropharyngeal secretions and swallowing frequency in predicting aspiration. *Dysphagia* 1996;11:99-103. <https://doi.org/10.1007/BF00417898>
- Molfenter SM, Steele CM. The relationship between residue and aspiration on the subsequent swallow: an application of the normalized residue ratio scale. *Dysphagia* 2013;28:494-500. <https://doi.org/10.1007/s00455-013-9459-8>
- Nordio S, Di Stadio A, Koch I, et al. Correlation between pharyngeal residue, penetration/aspiration and nutritional modality: a cross-sectional study in patients with neurogenic dysphagia. *Acta Otorhinolaryngol Ital* 2020;40:38-43. <https://doi.org/10.14639/0392-100X-2136>
- Langmore SE. History of fiberoptic endoscopic evaluation of swallowing for evaluation and management of pharyngeal dysphagia: changes over the years. *Dysphagia* 2017;32:27-38. <https://doi.org/10.1007/s00455-016-9775-x>
- Langmore SE, Schatz K, Olsen N. Endoscopic and videofluoroscopic evaluations of swallowing and aspiration. *Ann Otol Rhinol Laryngol* 1991;100:678-681. <https://doi.org/10.1177/000348949110000815>
- Pisegna JM, Langmore SE. Parameters of instrumental swallowing evaluations: describing a diagnostic dilemma. *Dysphagia* 2016;31:462-472. <https://doi.org/10.1007/s00455-016-9700-3>
- Kelly AM, Leslie P, Beale T, et al. Fiberoptic endoscopic evaluation of swallowing and videofluoroscopy: does examination type influence perception of pharyngeal residue severity? *Clin Otolaryngol* 2006;31:425-432. <https://doi.org/10.1111/j.1749-4486.2006.01292.x>
- Pearson WG, Molfenter SM, Smith ZM, et al. Image-based measurement of post-swallow residue: the normalized residue ratio scale. *Dysphagia* 2013;28:167-177. <https://doi.org/10.1007/s00455-012-9426-9>
- Espalier F, Fanous A, Aviv J, et al. International consensus (ICON) on assessment of oropharyngeal dysphagia. *Eur Ann Otorhinolaryngol Head and Neck Dis* 2018;135:S17-S21. <https://doi.org/10.1016/j.anorl.2017.12.009>
- Farneti D. Pooling score: an endoscopic model for evaluating severity of dysphagia. *Acta Otorhinolaryngol Ital* 2008;28:135-140.

- ¹⁴ Neubauer PD, Hersey DP, Leder SB. Pharyngeal residue severity rating scale based on fiberoptic endoscopic evaluation of swallowing: a systematic review. *Dysphagia* 2016;31:352-359. <https://doi.org/10.1007/s00455-015-9682-6>
- ¹⁵ Neubauer PD, Rademaker AW, Leder SB. The Yale pharyngeal residue severity rating scale: an anatomically defined and image-based tool. *Dysphagia* 2015;30:521-528. <https://doi.org/10.1007/s00455-015-9631-4>
- ¹⁶ Gerschke M, Schöttker-Königer T, Förster A, et al. Validation of the German version of the Yale pharyngeal residue severity rating scale. *Dysphagia* 2019;34:308-314. <https://doi.org/10.1007/s00455-018-9935-2>
- ¹⁷ Atar Y, Atar S, Ilgin C, et al. Validity and reliability of the Turkish translation of the Yale Pharyngeal Residue Severity rating scale. *Dysphagia* 2022;37:655-663. <https://doi.org/10.1007/s00455-021-10316-1>
- ¹⁸ Guillemin F. Cross-cultural adaptation and validation of health status measures. *Scand J Rheumatol* 1995;24:61-63. <https://doi.org/10.3109/03009749509099285>
- ¹⁹ Gjersing L, Caplehorn JR, Clausen T. Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. *BMC Med Res Methodol* 2010;10:13. <https://doi.org/10.1186/1471-2288-10-13>
- ²⁰ Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract* 2011;17:268-274. <https://doi.org/10.1111/j.1365-2753.2010.01434.x>
- ²¹ Cichero JAY, Lam PTL, Chen J, et al. Release of updated International Dysphagia Diet Standardisation Initiative Framework (IDDSI 2.0). *J Texture Stud* 2020;51:195-196. <https://doi.org/10.1111/jtxs.12481>
- ²² Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-745. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
- ²³ Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
- ²⁴ Fleiss JL, Levin B, Paik MC. *Statistical methods for rates and proportions*. Third Edition. Hoboken, NJ: Wiley; 2003.
- ²⁵ Rocca S, Pizzorni N, Valenza N, et al. Reliability and construct validity of the Yale Pharyngeal Residue Severity rating scale: performance on videos and effect of bolus consistency. *Diagnostics (Basel)* 2022;12:1897. <https://doi.org/10.3390/diagnostics12081897>

Appendix I

English version of the Yale Pharyngeal Residue Severity Rating Scale (YPRSRS).

Definitions for severity of vallecula residue			
I	None	0%	No residue
II	Trace	1-5%	Trace coating of the mucosa
III	Mild	5-25%	Epiglottic ligament visible
IV	Moderate	25-50%	Epiglottic ligament covered
V	Severe	> 50%	Filled to epiglottic rim

Definitions for severity of pyriform sinus residue			
I	None	0%	No residue
II	Trace	1-5%	Trace coating of mucosa
III	Mild	5-25%	Up wall to quarter full
IV	Moderate	25-50%	Up wall to half full
V	Severe	> 50%	Filled to aryepiglottic fold

Italian version of the Yale Pharyngeal Residue Severity Rating Scale (IT-YPRSRS).

Definitions for severity of vallecula residue			
I	Nessuno	0%	Assenza di residui
II	Tracce	1-5%	Tracce che rivestono la mucosa
III	Lieve	5-25%	Legamento glosso-epiglottico visibile
IV	Moderato	25-50%	Legamento glosso-epiglottico coperto
V	Grave	> 50%	Riempimento fino al margine libero dell'epiglottide

Definitions for severity of pyriform sinus residue			
I	Nessuno	0%	Assenza di residui
II	Tracce	1-5%	Tracce che rivestono la mucosa
III	Lieve	5-25%	Riempimento fino a un quarto della parete del seno
IV	Moderato	25-50%	Riempimento fino a metà della parete del seno
V	Grave	> 50%	Riempimento fino alla plica ariepiglottica