Visual Analysis of Swallowing Efficiency and Safety (VASES): Establishing Criterion-Referenced Validity and Concurrent Validity

James A. Curtis[1]
James C. Borders[1]
Michelle S. Troche[1]

[1]Laboratory for the Study of Upper Airway Dysfunction
Communication Sciences and Disorders Program
Department of Biobehavioral Sciences
Teachers College, Columbia University, New York, NY

Corresponding Author:
James A. Curtis, PhD, CCC-SLP, BCS-S
Email: jac2406@tc. columbia.edu

Abstract

Purpose: The primary aim of this study was to examine the criterion-referenced validity of the Visual Analysis of Swallowing Efficiency and Safety (VASES). As a secondary aim, we examined the concurrent validity of using verbal numerical ratings for VASES as a potential substitute for visual analogue scale ratings.

Method: Fifty-seven novice raters were prospectively recruited to rate 26 FEES images (2x each, randomized) – once using VASES and once using a criterion-referenced scale. Ratings were made for the valleculae, piriforms, epiglottis, laryngeal vestibule, vocal folds, and subglottis. Criterion validity was determined by examining the correlation between VASES and the criterion-referenced scales. The novice raters also provided visual analogue scale ratings following verbal numerical ratings. Concurrent validity of using verbal numerical ratings as a potential substitute for visual analogue scale ratings was determined by examining the correlation and absolute agreement between both rating methods.

Results: 3,587 ratings were analyzed. Spearman's correlation revealed strong correlations between VASES ratings and criterion-referenced ratings across all anatomic landmarks (ρ = .882-.915). Lin's concordance revealed substantial agreement between numerical ratings and visual analogue scale ratings (ρ_c = .986).

Conclusions: The strong correlations between VASES and the criterion-referenced scales suggest that VASES is a valid method for interpreting pharyngeal residue, penetration, and aspiration during FEES. Furthermore, numerical ratings exhibited substantial agreement with visual analogue scales. This suggests that clinicians could provide verbal numerical ratings in lieu of visual analogue scale ratings as a potential way to enhance the ease and feasibility of implementing VASES into clinical practice.

Introduction

VASES

Visual Analysis of Swallowing Efficiency and Safety (VASES) is a newly developed rating framework used to judge pharyngeal residue, penetration, and aspiration during flexible endoscopic evaluations of swallowing (FEES) (J. A. Curtis et al., 2021). VASES uses 100-point visual analogue scales, clearly defined anatomic and temporal boundaries, and additional "secondary rules" all intended to improve the standardization and transparency of rating functional swallowing outcomes during FEES. Visual-perceptual assessments are used during VASES to estimate the amount of residue "filling" the oropharynx-valleculae and hypopharynx-piriforms, and "covering" the epiglottis, laryngeal vestibule, vocal folds, and subglottis. A 100-point visual analogue scale is used to complete the residue ratings. The visual analogue scale contains two verbal anchors: "0% (no filling/covering)", located at the left-most point of the scale, and "100% (completely filled/covered)", at the right-most endpoint of the scale. No other verbal-numerical anchors are provided on the scale.

Validating FEES Rating Scales

The reliability and validity of rating scales used for FEES should be well-understood prior to implementation in clinical and research practices. Preliminary data demonstrates that VASES yields adequate intra- and inter-rater reliability amongst novice clinicians (Curtis et al., 2021). However, the validity of VASES has not yet been established. Criterion validity is established by comparing a new scale (i.e., VASES in this case) to other previously validated criterion-referenced scales (Mokkink et al., 2010). Two such validated FEES rating scales include the Yale Pharyngeal Residue Severity Rating Scale (YPRSRS) and the Boston Residue and Clearance Scale (BRACS).

The YPRSRS is an anatomically defined imaged-based scale intended to estimate the amount of residue contained within the valleculae and piriforms using a 5-point ordinal rating scale (Neubauer et al., 2015). The five severity levels include none (0% filling), trace (1-5% filling), mild (5-25% filling), moderate (25-50% filling), severe (>50% filling). The YPRSRS is an "anatomically defined" scale because it uses anatomic descriptors (e.g., epiglottic ligament visible) to categorize residue severity levels, and is "image-based" because it includes exemplar images for each severity level and anatomic landmark.

The BRACS is an anatomically defined scale intended to estimate the amount of residue within the pharynx and larynx using a 4-point ordinal rating scale (Kaneoka et al., 2014). The four severity

levels include none/coating, mild (<1/3 covering/filling), moderate (1/3-2/3 covering/filling), severe (>2/3 covering/filling). Severity ratings are applied separately across 12 anatomic landmarks within the pharynx and larynx. In addition to rating residue, the BRACS can be used to produce a sum score which takes into account global severity of residue and clearing ability.

Both the YPRSRS and BRACS use ordinal, categorical methods to rate residue. Despite this, emerging research supports using continuous interval-based scales to rate residue (Pisegna et al., 2020; Pisegna, Borders, et al., 2018; Pisegna, Kaneoka, et al., 2018; Steele et al., 2020). For example, research by Dr. Pisegna and colleagues has found 100-point visual analogue scales facilitate greater precision in pharyngeal residue ratings when compared to traditional categorical ratings (Pisegna, Kaneoka, et al., 2018). It is for this reason that visual analogue scales were incorporated into the original development of VASES.

Visual Analogue Scales and Verbal Numerical Rating Scales

Visual analogue scales can be useful for rating residue (Pisegna, Kaneoka, et al., 2018) and have also been shown to be efficacious in other areas of speech pathology including auditory-perceptual assessment of voice, resonance, and speech (Bettens et al., 2018; Castick et al., 2017; Kempster et al., 2009; San Segundo & Skarnitzl, 2019; Sussman & Tjaden, 2012; Tjaden et al., 2014; Zraick et al., 2011). Visual analogue scales require measuring the location of a mark on a 100-point line, either digitally or using pen and paper. However, visual analogue scales limit the ability to verbally describe impairment to colleagues, which decreases the ease and feasibility of integrating visual analogue scales in clinical practice. Given that VASES uses visual analogue scales to estimate the numerical percentage (%) of residue filling or covering an anatomic landmark, rather than a subjective impression of "perceived severity", it stands to reason that verbal numerical ratings (i.e., simply selecting a number 0 through 100) could be a potential substitute for visual analogue scale ratings. Research comparing visual analogue scales and verbal numerical scales to estimate pain and pruritus have found a strong correlation and high level of agreement between the two rating methods (Adam et al., 2012; Hjermstad et al., 2011; Holdgate et al., 2003; Hollen et al., 2005; Mohan et al., 2010; Reich et al., 2016). If verbal numerical ratings exhibit high agreement (concurrent validity) with visual analogue scales, then verbal numerical ratings may be a potential substitute for visual analogue scales for VASES. This substitute could potentially enhance the feasibility of implementing VASES into clinical practice.

Aims

The primary aim of this study was to determine the criterion validity of VASES by comparing the method used to rate residue during VASES compared to two previously validated FEES ratings scales: YPRSRS and BRACS. These scales were chosen because, for scales describing pharyngeal residue, YPRSRS exhibits the strongest level of validity, and for scales describing laryngeal residue, BRACS exhibits the strongest level validity (Neubauer et al., 2016; Swan et al., 2018). We hypothesized there would be a strong correlation between VASES with the criterion-referenced scales. The secondary aim of this study was to determine the concurrent validity of verbal numerical ratings and visual analogue scale ratings. We hypothesized that verbal numerical ratings would exhibit substantial agreement with visual analogue scales, thus providing evidence that they could be used for rating residue for VASES.

Methods

Residue Rating Image Selection

The study was approved by the university's Institutional Review Board (IRB #: 21-071). Two expert judges (JC and JB) reviewed records of 250 FEES. The two expert judges for this study convened to identify one endoscopic image associated with each severity level for each VASES anatomic landmark using either the YPRSRS (valleculae and piriforms) and BRACS (epiglottis, laryngeal vestibule, vocal folds, subglottis). The severity level for each residue endoscopic image was agreed upon by the expert judges and determined to meet the criteria of the YPRSRS and BRACS (see "Severity Across Anatomic Landmarks" in the supplemental materials for the images provided for the raters).

The FEES video clips were pulled from an outpatient clinical research database of people with dysphagia and neurodegenerative disease. The FEES equipment used in these video clips was a 3.0 mm diameter flexible distal chip laryngoscope (ENT-5000; Cogentix Medical, New York, USA) and video system with integrated LED light source LCD display (Cogentix Medical, DPU-7000A). During the FEES, the flexible laryngoscope was passed transnasally, without the use of topical anesthetic or vasoconstrictors. The tip of the endoscope was positioned within the oropharynx in order to visualize the pharynx, larynx, and subglottis before, during, and after all swallows. As needed, the endoscope was advanced throughout the pharynx and laryngeal vestibule after each swallow to more closely

inspect residue patterns throughout the pharynx, laryngeal and subglottic spaces. Boluses included in the FEES included thin liquid, mildly thick liquid, puree, and dry solids. All liquid boluses were artificially colored with either blue dye, green dye, white dye, barium, or a combination these colorants (Curtis et al., 2019; Curtis et al., 2020). FEES were completed by, or under the direct supervision of, a speech-language pathologist experienced in the performance and interpretation of FEES.

Procedure

Participants

Fifty-seven raters were recruited from a graduate school speech-language pathology program. All raters were master-level students enrolled in one of two sections of a dysphagia class at the time of the study. The raters were in the second semester of their training program at the time of the study without any prior internship training experiences. This study was completed virtually, in real-time, on the student's personal computers, as part of a FEES interpretation training.

Criterion Validity: Residue Ratings

Brief tutorials on how to rate residue using the YPRSRS, BRACS, and VASES were presented to the novice raters using PowerPoint immediately prior to starting the ratings for this study. For YPRSRS, the tutorial included displaying the exemplar images presented in figures 1 and 2 of the original YPRSRS manuscript. Additionally, the definition of each severity category was outlined, including the severity rating (none, trace, mild, moderate, severe), the percentages associated with each severity rating (0 %, 1-5 %, 5-25 %, 25-50 %, >50 %), and the anatomically defined verbal descriptors associated with each severity rating (e.g., "epiglottic ligament visible" for the "mild" severity rating). For BRACS, the tutorial included outlining the scale used to define the amount of residue seen endoscopically, including: none/coating; mild = covering/filling of <1/3 of the location; moderate = covering/filling of 1/3-2/3 of the location; severe = covering/filling of >2/3 of the location. For VASES, the tutorial included how to create transparent, standardized anatomic boundaries for the oropharynx, hypopharynx, epiglottis, laryngeal vestibule, vocal folds, and subglottis. The tutorial explicitly stated that the raters should: "estimate the amount of (oropharyngeal or hypopharyngeal) residue filling the (valleculae or piriforms)", "estimate the amount of (epiglottic, laryngeal vestibule, or vocal fold) surface area covered by residue", or "estimate the amount of subglottic shelf surface area covered by all subglottic residue" depending on the anatomic landmark being rated. For example, a

rating of "100" for the subglottis would indicate that 100% of the subglottic shelf surface area is covered with subglottic residue (i.e., residue from the subglottic shelf, cricoid cartilage, and trachea).

Raters were informed that they would rate the amount of residue seen on just one pre-specified anatomic landmark, using just one of the residue rating scales, across 52 different endoscopic images. All endoscopic still images were displayed using PowerPoint via an online video conferencing platform (Zoom Video Communications, Inc.). The endoscopic images presented to the novice raters were still images from previously recorded FEES, typically taken after the initial swallow of a bolus trial. Each image was presented for approximately 30 seconds before being removed and presented with a new still image. All residue ratings were uploaded directly into REDCap.

Residue ratings of the endoscopic images were made for six anatomic landmarks: oropharynx-valleculae, hypopharynx-piriforms, epiglottis, laryngeal vestibule, vocal folds, subglottis. These landmarks were used since they represent the six residue ratings used for VASES. Within each anatomic landmark, one endoscopic image associated with each severity level for the criterion-referenced scale was presented. Each endoscopic image was rated twice, once using VASES and once using the criterion-referenced scale. The criterion-referenced scales included the YPRSRS for the valleculae and piriforms and the BRACS for the epiglottis, laryngeal vestibule, and vocal folds. No FEES rating scale currently exists which quantifies subglottic residue (aspiration amount), and therefore, no criterion-reference scale was used for this anatomic landmark. However, raters were asked to rate subglottic residue using VASES and using the same 4-point categorical rating method used in BRACS (none, <1/3, 1/3-2/3, >2/3) in order to examine the relationship between categorical and continuous rating methods for subglottic residue. Therefore, raters rated endoscopic images once using VASES and once using YPRSRS for the valleculae and piriforms, and once using VASES and once using BRACS for the epiglottis, laryngeal vestibule, vocal folds, and subglottis.

For the YPRSRS and BRACS, definitions for each severity were provided on each PowerPoint slide for raters to use as a referent guide during rating. For YPRSRS only, accompanying exemplar images were provided. Similarly, the rating rules for VASES and a picture of the anatomic boundaries (Figure 1) were provided on each PowerPoint slide for raters to use as a referent guide during rating. For VASES, a digital visual analogue scale which contained verbal anchors of "0% (none)" and "100% (complete)" on the left and right endpoints of the scale was provided. A digital marker was present halfway along the line which raters subsequently moved according to how much residue they perceived – the raters were never told that the starting point of the digital marker represented 50/100

on the scale. No other verbal descriptors or dash marks were present on the visual analogue scale line (Figure 2). Residue rating were completed sequentially by each anatomic landmark, but with scale types and severity levels randomized within each landmark grouping.

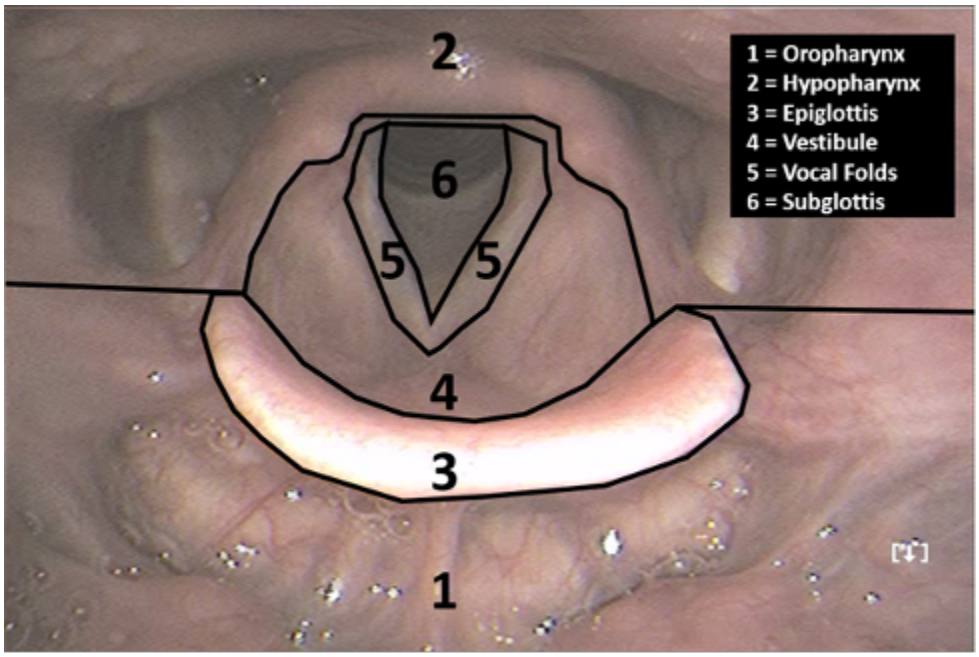


Figure 1: Picture of the anatomic landmarks provided during pre- and post-training

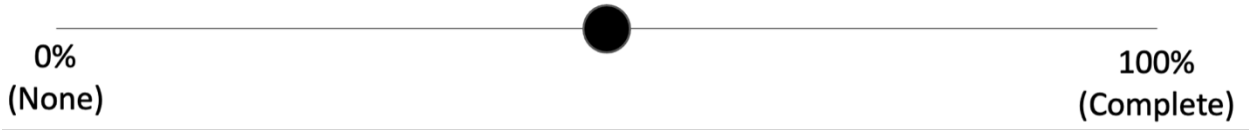


Figure 2: Example of the visual analogue scale range from 0% (none) to 100% (complete). The central black point (set currently to 50/100) represents the digital mark that was moved along the scale.

Concurrent Validity: Numerical Ratings

Following completion of the residue ratings for the six anatomic landmarks, novice raters were presented with a series of 12 additional PowerPoint slides. Each slide contained the following prompt: "Using VASES, rate where you think [numerical rating] is on the visual analogue scale." Each PowerPoint slide contained one of the following numerical ratings (randomized): 0, 3, 5, 15, 25, 33,

38, 50, 66, 75, 83, 100. These numbers were used because they represent the middle or outer most boundaries associated with each severity level for YPRSRS and BRACS. Raters used the same visual analogue scale that was used for VASES, which contained only the two end-point verbal anchors, but no other descriptors or dash marks on the line.

Statistical Analysis

All analyses were performed in R version 4.0.3 (R Core Team, 2019). A familywise alpha was set at < .05, and Holm-Bonferroni adjustments were used to correct for multiple comparisons.

Criterion validity was statistically analyzed using Spearman's correlation by comparing VASES to YPRSRS for oropharynx/valleculae and hypopharynx/piriforms residue ratings. Criterion validity was also statistically analyzed using Spearman's correlation by comparing VASES to BRACS for the epiglottis, laryngeal vestibule, and vocal folds residue ratings. Because no validated FEES rating scale exists quantifying subglottic residue (aspiration amount), the BRACS 4-point categorical rating method was applied to the subglottis and compared to VASES ratings. Correlations were considered weak if $\rho <$ | 0.4 |, moderate if | 0.7 | $< \rho \geq$ | 0.4 |, strong if | 1.0 | $< \rho \geq$ | 0.7 |, and perfect if $\rho =$ | 1.0 |. A correlation of $\rho \geq .7$ was set a priori as the cut-off value for determining if VASES was a considered to be valid for rating residue for each anatomic landmark.

Concurrent validity was statistically analyzed by comparing verbal numerical ratings with visual analogue scale ratings. Because both are measured using a 100-point continuous scale, Lin's concordance correlation coefficient was used to examine concurrent validity of verbal numerical ratings since both verbal numerical ratings and the visual analogue scale. Strength-of-agreement was considered poor if $\rho_c < 0.90$, moderate if $\rho_c = 0.90$ to 0.95, substantial if $\rho_c = 0.95$ to 0.99, and almost perfect if $\rho_c > .99$ (Lin, 1989; McBride, 2005; Steichen & Cox, 2002). An agreement of $\rho_c \geq .95$ was set a priori as the cut-off value for determining if verbal numerical ratings were a valid substitution for visual analogue scales when rating residue during VASES.

Results

Criterion Validity: Comparing VASES to YPRSRS and BRACS

A total of 57 novice raters were recruited, yielding an analysis of 2,964 criterion-referenced validity ratings. Spearman's correlation revealed strong, significant correlations between VASES, YPRSRS, and BRACS for all anatomic landmarks. Specifically, Spearman's correlation was $\rho = .884$,

$p < .0005$ between VASES and YPRSRS for the oropharynx/valleculae (Figure 3) and $\rho = .893$, $p < .0005$ for the hypopharynx/piriforms (Figure 4). Additionally, Spearman's correlation was $\rho = .895$, $p < .0005$ between VASES and BRACS for the epiglottis (Figure 5), $\rho = .915$, $p < .0005$ for the laryngeal vestibule (Figure 6), $\rho = .898$, $p < .0005$ for the vocal folds (Figure 7), and $\rho = .882$, $p < .0005$ for the subglottis (Figure 8).

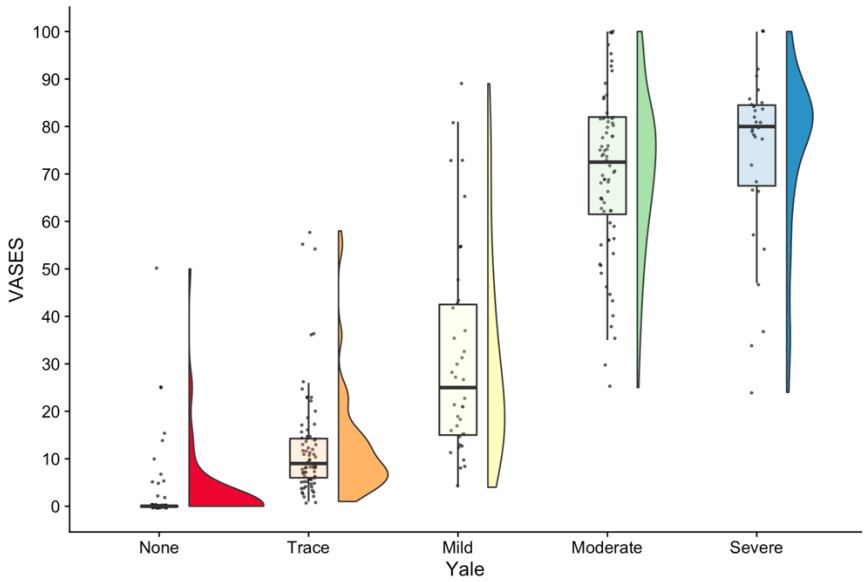


Figure 3: Correlation between VASES and YPRSRS for the oropharynx-valleculae

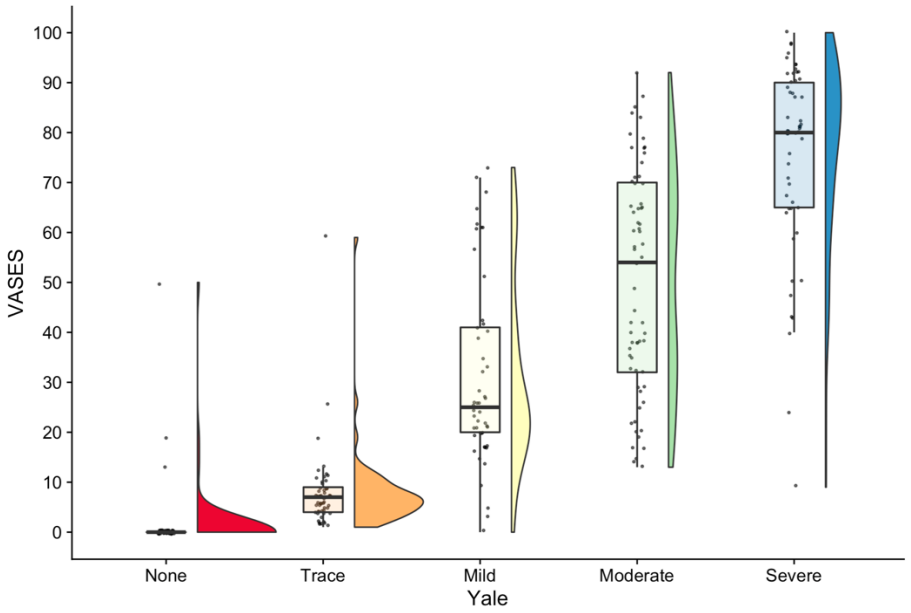


Figure 4: Correlation between VASES and YPRSRS for the hypopharynx-piriforms

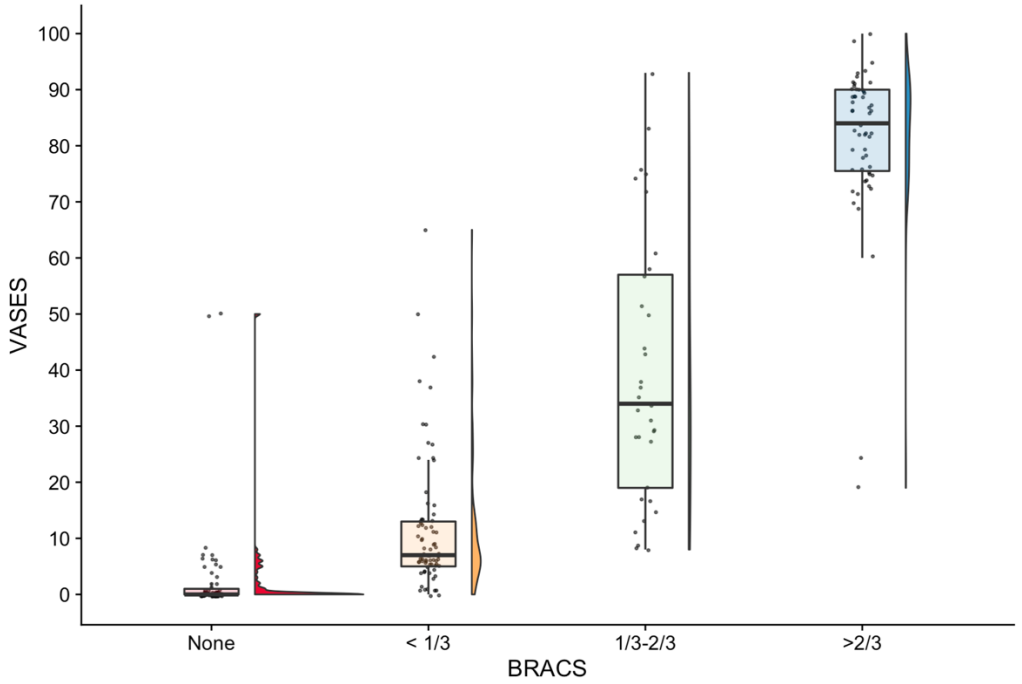


Figure 5: Correlation between VASES and BRACS for the epiglottis

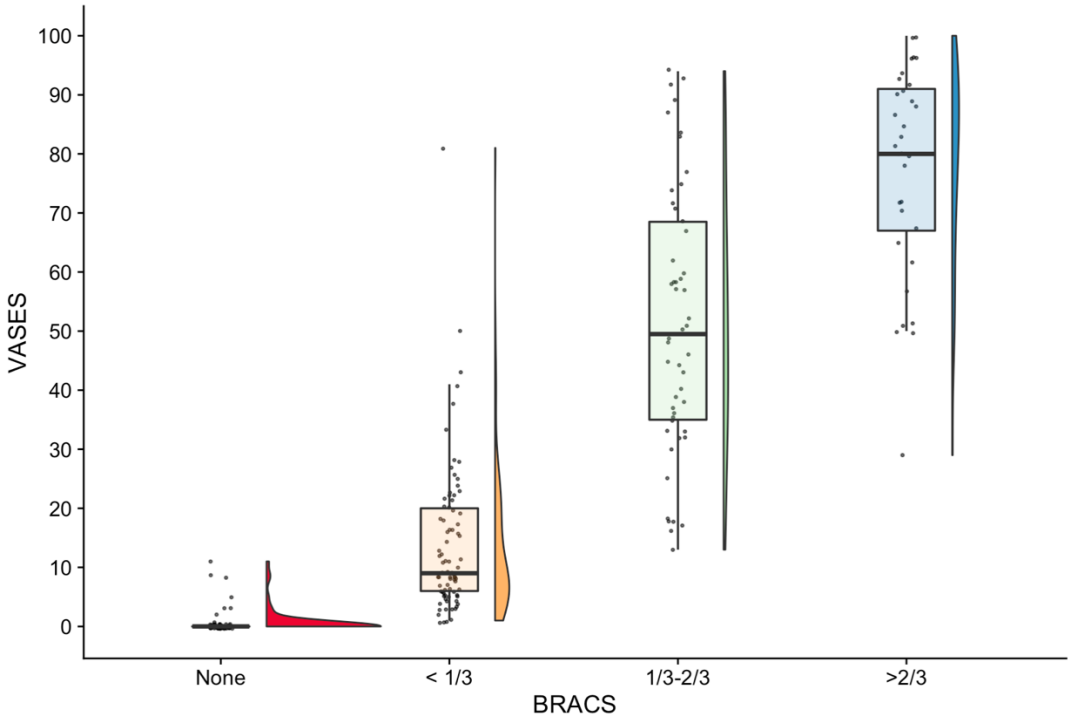


Figure 6: Correlation between VASES and BRACS for the laryngeal vestibule

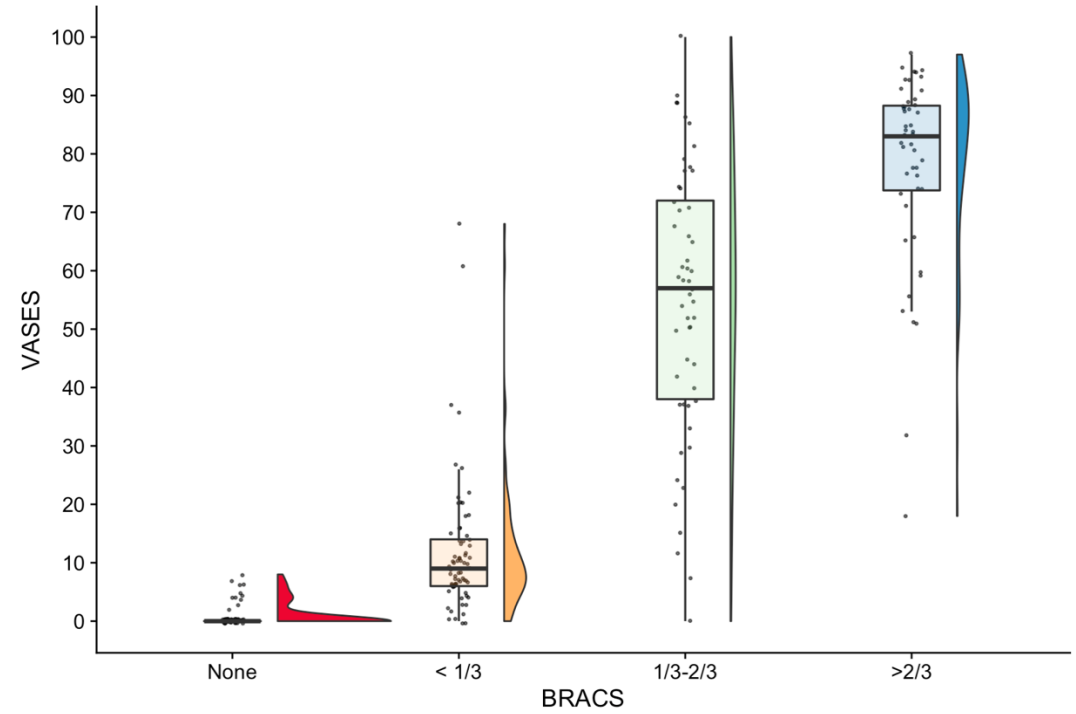


Figure 7: Correlation between VASES and BRACS for the vocal folds

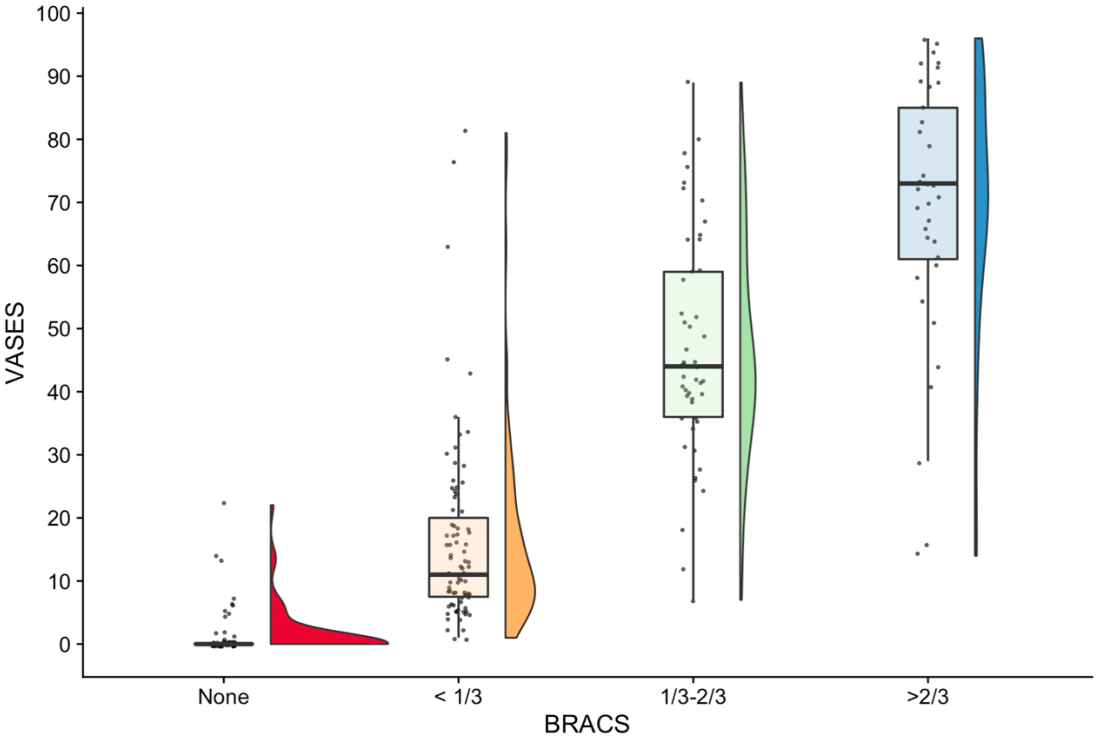


Figure 8: Correlation between VASES and BRACS for the subglottis

Concurrent Validity: Comparing Verbal Numerical Ratings and Visual Analogue Scale Ratings

All 57 novice raters completed these ratings as well, yielding an analysis of 684 concurrent validity ratings. Descriptive statistics of the visual analogue scale ratings for each verbal numerical rating are outlined in Table 1. For verbal numerical prompts ≤ 15 (excluding 0), visual analogue scales were an average 1.6 points greater than the verbal numerical prompt. For verbal numerical prompts ≥ 25 (excluding 100), visual analogue scales were an average 3.7 points less than the verbal numerical prompt. Lin's concordance correlation revealed substantial agreement between the verbal numerical ratings (prompts) and the visual analogue scale ratings, $\rho_c = .986$ (95% CI: .984 - .988). There was a scale shift of $\omega = 1.034$, a location shift of $v = 0.056$, and an accuracy of $\chi_a = .997$ (Figure 9).

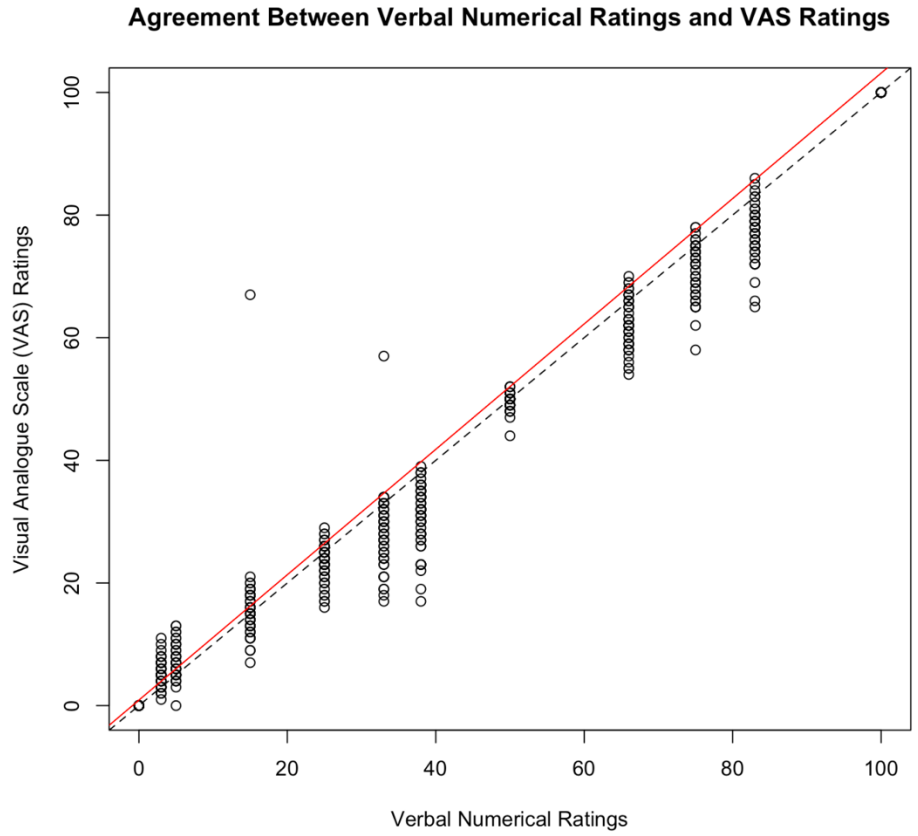


Figure 9: Agreement between verbal numerical ratings (prompts) and visual analogue scale (VAS) ratings, with 57 data points per column. A perfect correlation is represented by the 45° dashed (black) line, while the line of best fit for these data is represented by the solid (red) line.

Table 1: Descriptive Statistics of Visual Analogue Scale Ratings Across Verbal Numerical Ratings

Verbal Numerical Rating	Visual Analogue Scale Ratings		
	Mean	SD	CV
0	0.0	0.0	0.0
3	5.0	2.1	0.4
5	6.9	2.6	0.4
15	16.0	7.5	0.5
25	23.4	2.8	0.1
33	28.5	5.9	0.2
38	31.2	4.9	0.2
50	49.7	1.3	0.0
66	62.3	3.9	0.1
75	71.2	4.1	0.1
83	77.2	4.2	0.1
100	100.0	0.0	0.0

Abbreviations: standard deviation (SD), coefficient of variation (CV)

Discussion

VASES is a newly established rating method intended to increase the standardization and transparency of measuring pharyngeal residue, penetration, and aspiration during FEES. It outlines specific anatomic and temporal boundaries, in addition to "secondary rules", to guide judgements of swallowing efficiency and safety as seen during FEES. Previous research has demonstrated that VASES facilitates good-to-excellent reliability amongst novice clinicians and is feasible to learn and implement into clinical practice. Results from this study build on prior research by establishing the validity of VASES for use in clinical and research practices.

Criterion validity for VASES was determined by examining its relationship with criterion-referenced scales – BRACS and YPRSRS. Visual inspection of the data in Figures 3-8 revealed a large spread of VASES ratings for moderate and severe categories for both YPRSRS and BRACS. However, these large spreads of data were created by outliers that do not reflect the majority of VASES ratings. Instead, the interquartile range (IQR), which depicts the 25th, 50th and 75th quartiles, should be used to visualize relationships between VASES, YPRSRS, and BRACS. When visually inspecting the IQRs for VASES, it was found that VASES ratings were contained largely within the numerical boundaries defined by YPRSRS and BRACS. For example, in Figure 8, the BRACS 1/3-2/3 severity category had

a VASES IQR extending from approximately 35 to 60. Furthermore, statistical analyses with Spearman correlations revealed strong, statistically significant relationships with VASES for both YPRSRS and BRACS. Together, these data support the use of VASES as a valid method to rate pharyngeal residue, penetration, and aspiration during FEES.

The YPRSRS and BRACS use 5- and 4-point ordinal rating scales to measure the amount of residue seen within the pharynx and larynx. While the YPRSRS provides anatomic-based descriptions of residue severity with accompanying exemplar images, neither the YPRSRS nor BRACS provide detailed descriptions on how to delineate the anatomic boundaries for the valleculae, piriforms, epiglottis, laryngeal vestibule, or vocal folds. Furthermore, neither of these scales provide methods of rating subglottic residue (aspiration amount). VASES was developed in part to address these gaps in FEES analysis. It uses a 100-point rating scale to judge the amount of residue filling the oropharynx, hypopharynx, epiglottis, laryngeal vestibule, vocal folds, and subglottis. Despite differences between the scales, including the use of different rating scale methods, having clearly defined vs. non-specific anatomic boundaries, and judging the entire oro- and hypopharynx rather than only the valleculae and piriforms, there was a strong correlation between these scales. These results demonstrate a high level of criterion-referenced validity for VASES and further support its valid use to judge pharyngeal residue during FEES.

However, in order for a scale to be widely adopted into clinical and research practices, it needs to be not just valid and reliable, but also feasible to implement. Our previous work demonstrates that VASES is feasible to learn and train. However, from an implementation standpoint, the present study sought to determine if verbal numerical ratings of 0-100 could be used as a valid substitute for the visual analogue scale in order to further increase clinical feasibility of VASES implementation. To do this, 12 numbers representing the middle and outer most boundaries of each severity level for the YPRSRS and BRACS were selected. Raters attempted to match the numbers with a visual analogue scale rating. However, it is noteworthy that there was an uneven distribution of numbers throughout the 100-point with seven numbers below 50 and four numbers above 50. This uneven distribution across the 100-point continuum may have resulted in the inability to detect areas of the visual analogue scale or numerical ratings that raters may generally avoid. This phenomenon, known as the “halo effect”, has been observed in similar research involving pharyngeal residue ratings (Pisegna et al., 2020). Despite this, results from the present study demonstrated substantial agreement and concurrent validity between verbal numerical ratings and visual analogue scales. This suggests verbal numerical

ratings can be confidently used as a valid substitute for visual analogue scales when rating the estimated amount of pharyngeal, laryngeal, and subglottic residue with VASES.

There are several limitations that should be considered when interpreting the results from this study. First, FEES images were used for scale validation rather than full-length video clips. This was done to ensure that the same residue from the same video frame was being assessed for both scales. However, by doing so, we were unable to determine how temporal boundaries (or lack thereof) inherent across rating scales may have impacted our results. Therefore, future studies should expand on the present findings by using full length video clips to compare if/how VASES differs from other validated scales when taking into account entire video clips. Second, anatomic boundaries are not clearly defined for YPRSRS or BRACS but are for VASES. Because all three scales were briefly taught prior to beginning ratings, it is possible that learning the anatomic boundaries for VASES may have influenced YPRSRS and BRACS ratings. Third, different colorants were randomly selected for the residue rating images. Whereas some images had colorants which elicited a coating effect, others were opaque but with no coating effect. While this should not affect the ability to compare VASES to the criterion-referenced scale within each category since the same image was used for both scales, this may limit the ability to compare ratings across severity categories. Lastly, data collected in this study were from novice raters in the first year of their speech-language pathology graduate training program. While current research suggests that experience does not significantly impact residue rating findings (Pisegna, Borders, et al., 2018), it is unknown how the findings from this study may have differed if using a group of people with a range of experience levels.

Conclusions

VASES is a newly established framework used to guide ratings of functional swallowing outcomes during FEES. It was developed to enhance the standardization, transparency, and reliability of FEES analysis. Results from this study demonstrate that the anatomic boundaries and rating methods used by VASES are highly correlated to criterion-referenced scales, demonstrating that VASES is also a valid method for FEES analysis. Furthermore, VASES may be rated with either 100-point visual analogue scales or 100-point verbal numerical ratings in order to improve the ease and feasibility of implementation into clinical practice.

References

Adam, R., Heisig, M., Phan, N. Q., Taneda, K., Takamori, K., Takeuchi, S., Furue, M., Blome, C., Augustin, M., Ständer, S., & Szepietowski, J. C. (2012). Visual analogue scale: Evaluation of the instrument for the assessment of pruritus. *Acta Dermato-Venereologica*, *92*(5), 497–501. https://doi.org/10.2340/00015555-1265

Bettens, K., Bruneel, L., Maryn, Y., De Bodt, M., Luyten, A., & Van Lierde, K. M. (2018). Perceptual evaluation of hypernasality, audible nasal airflow and speech understandability using ordinal and visual analogue scaling and their relation with nasalance scores. *Journal of Communication Disorders*, *76*, 11–20. https://doi.org/10.1016/j.jcomdis.2018.07.002

Castick, S., Knight, R. A., & Sell, D. (2017). Perceptual judgments of resonance, nasal airflow, understandability, and acceptability in speakers with cleft palate: Ordinal versus visual analogue scaling. *Cleft Palate-Craniofacial Journal*, *54*(1), 19–31. https://doi.org/10.1597/15-164

Curtis, J. A., Borders, J. C., Perry, S. E., Dakin, A. E., Seikaly, Z. N., & Troche, M. S. (2021). Visual Analysis of Swallowing Efficiency and Safety (VASES): A standardized approach to rating pharyngeal residue, penetration, and aspiration during FEES. *Dysphagia*, 1–19. https://doi.org/10.1007/s00455-021-10293-5

Curtis, J. A., Seikaly, Z. N., Dakin, A. E., & Troche, M. S. (2020). Detection of aspiration, penetration, and pharyngeal residue during flexible endoscopic evaluation of swallowing (FEES): Comparing the effects of color, coating, and opacity. *Dysphagia*. https://doi.org/10.1007/s00455-020-10131-0

Curtis, J., Perry, S., & Troche, M. S. (2019). Detection of airway invasion during flexible endoscopic evaluations of swallowing: Comparing barium, blue dye, and green dye. *American Journal of Speech-Language Pathology*, *28*(2), 515–520. https://doi.org/10.1044/2018_AJSLP-18-0119

Hjermstad, M. J., Fayers, P. M., Haugen, D. F., Caraceni, A., Hanks, G. W., Loge, J. H., Fainsinger, R., Aass, N., & Kaasa, S. (2011). Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: A systematic literature review. *Journal of Pain and Symptom Management*, *41*(6), 1073–1093. https://doi.org/10.1016/j.jpainsymman.2010.08.016

Holdgate, A., Asha, S., Craig, J., & Thompson, J. (2003). Comparison of a verbal numeric rating scale with the visual analogue scale for the measurement of acute pain. *Emergency Medicine*, *15*, 441–446.

Hollen, P. J., Gralla, R. J., Kris, M. G., Mccoy, S., Donaldson, G. W., & Moinpour, C. M. (2005). A

comparison of visual analogue and numerical rating scale formats for the Lung Cancer Symptom Scale (LCSS): Does format affect patient ratings of symptoms and quality of life? *Quality of Life Research*, *14*(3), 837–847.

Kaneoka, A. S., Langmore, S. E., Krisciunas, G. P., Field, K., Scheel, R., Mcnally, E., Walsh, M. J., O'Dea, M. B., & Cabral, H. (2014). The Boston residue and clearance scale: Preliminary reliability and validity testing. *Folia Phoniatrica et Logopaedica*, *65*(6), 312–317. https://doi.org/10.1159/000365006

Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*, *18*, 124–132. https://doi.org/10.1044/1058-0360(2008/08-0017)

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*(1), 255–268. https://www.jstor.org/stable/2532051

McBride, G. B. (2005). A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. *NIWA Client Report: HAM2005-062*, 1–10.

Mohan, H., Ryan, J., Whelan, B., & Wakai, A. (2010). The end of the line? The Visual Analogue Scale and Verbal Numerical Rating Scale as pain assessment tools in the emergency department. *Emergency Medicine Journal*, *27*(5), 372–375. https://doi.org/10.1136/emj.2007.048611

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Neubauer, P. D., Hersey, D. P., & Leder, S. B. (2016). Pharyngeal Residue Severity Rating Scales Based on Fiberoptic Endoscopic Evaluation of Swallowing: A Systematic Review. *Dysphagia*, *31*(3), 352–359. https://doi.org/10.1007/s00455-015-9682-6

Neubauer, P. D., Rademaker, A. W., & Leder, S. B. (2015). The Yale Pharyngeal Residue Severity Rating Scale: An anatomically defined and image-based tool. *Dysphagia*, *30*(5), 521–528. https://doi.org/10.1007/s00455-015-9631-4

Pisegna, J. M., Borders, J. C., Kaneoka, A., Coster, W. J., Leonard, R., & Langmore, S. E. (2018). Reliability of untrained and experienced raters on FEES: rating overall residue is a simple task.

Dysphagia, *33*(5), 645–654. https://doi.org/10.1007/s00455-018-9883-x

Pisegna, J. M., Kaneoka, A., Coster, W. J., Leonard, R., & Langmore, S. E. (2020). Residue ratings on FEES: trends for clinical application of residue measurement. *Dysphagia*, 1–8. https://doi.org/10.1007/s00455-019-10089-8

Pisegna, J. M., Kaneoka, A., Leonard, R., & Langmore, S. E. (2018). Rethinking residue: determining the perceptual continuum of residue on FEES to enable better measurement. *Dysphagia*, *33*(1), 100–108. https://doi.org/10.1007/s00455-017-9838-7

R Core Team. (2019). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing* (pp. 115–124). https://doi.org/10.1016/j.dendro.2008.01.002

Reich, A., Riepe, C., Anastasiadou, Z., Medrek, K., Augustin, M., Szepietowski, J. C., & Ständer, S. (2016). Itch assessment with visual analogue scale and numerical rating scale: Determination of minimal clinically important difference in chronic itch. *Acta Dermato-Venereologica*, *96*(7), 978–980. https://doi.org/10.2340/00015555-2433

San Segundo, E., & Skarnitzl, R. (2019). A computer-based tool for the assessment of voice quality through visual analogue scales: VAS-simplified vocal profile analysis. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2019.10.007

Steele, C. M., Peladeau-Pigeon, M., Nagy, A., & Waito, A. A. (2020). Measurement of pharyngeal residue from lateral view videofluoroscopic images. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1404–1415. https://doi.org/10.1044/2020_JSLHR-19-00314

Steichen, T. J., & Cox, N. J. (2002). A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. *The Stata Journal*, *2*(2), 183–189. https://doi.org/https://doi.org/10.1177/1536867X0200200206

Sussman, J. E., & Tjaden, K. (2012). Perceptual measures of speech from individuals with Parkinson's disease and multiple Sclerosis: Intelligibility and beyond. *Journal of Speech, Language, and Hearing Research*, *55*(4), 1208–1219. https://doi.org/10.1044/1092-4388(2011/11-0048)

Swan, K., Cordier, R., Brown, T., & Speyer, R. (2018). Psychometric properties of visuoperceptual measures of videofluoroscopic and Fibre-Endoscopic Evaluations of Swallowing: a systematic review. *Dysphagia*, 1–32. https://doi.org/10.1007/s00455-018-9918-3

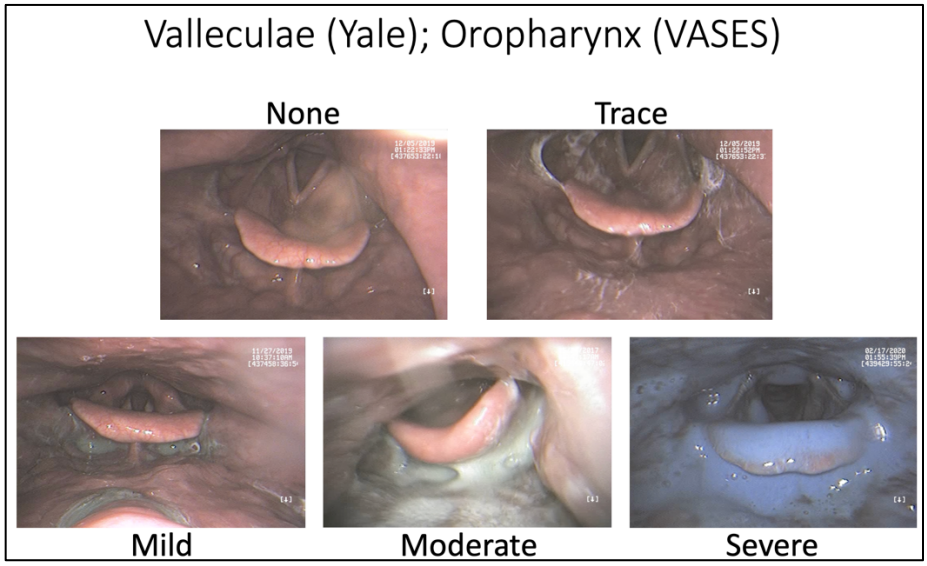
Tjaden, K., Sussman, J. E., & Wilding, G. E. (2014). Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in parkinson's disease and multiple sclerosis. *Journal of Speech, Language, and Hearing Research*, *57*(3), 779–792. https://doi.org/10.1044/2014_JSLHR-S-12-

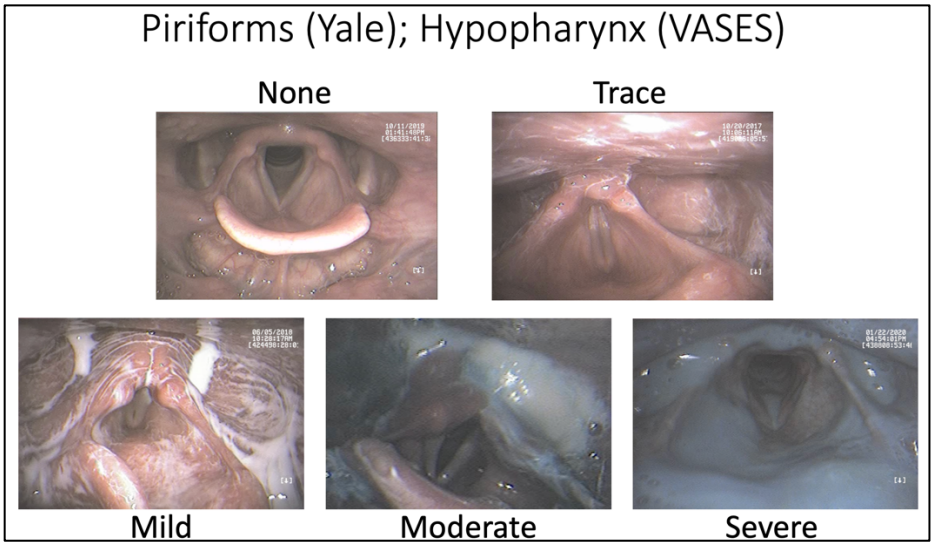
0372

Zraick, R. I., Kempster, G. B., Connor, N. P., Thibeault, S., Klaben, B. K., Bursac, Z., Thrush, C. R., & Glaze, L. E. (2011). Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *American Journal of Speech-Language Pathology*, *20*(1), 14–22. https://doi.org/10.1044/1058-0360(2010/09-0105)

Supplemental Material

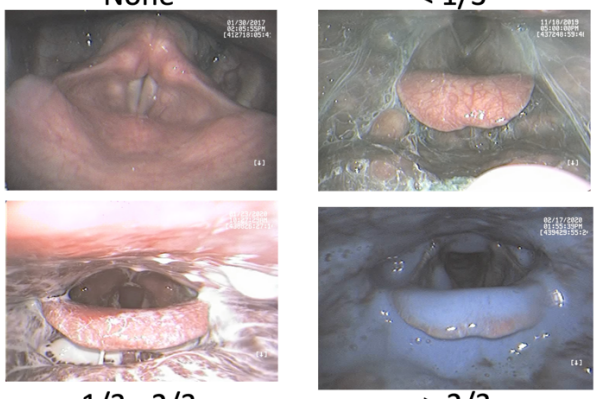
This supplemental document contains the endoscopic images that were provided to the raters for rating of the YPRSRS, BRACS, and VASES.





Epiglottis (BRACS & VASES)

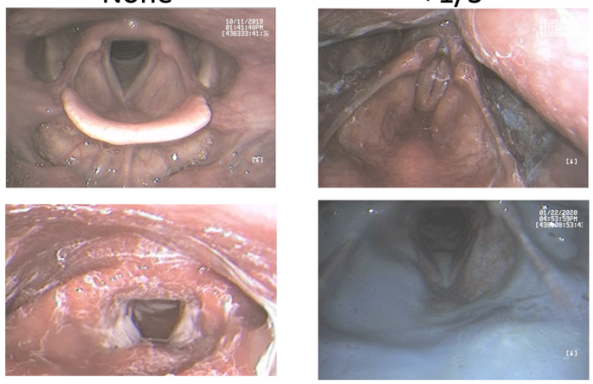
None | < 1/3



1/3 - 2/3 | > 2/3

Laryngeal Vestibule (BRACS & VASES)

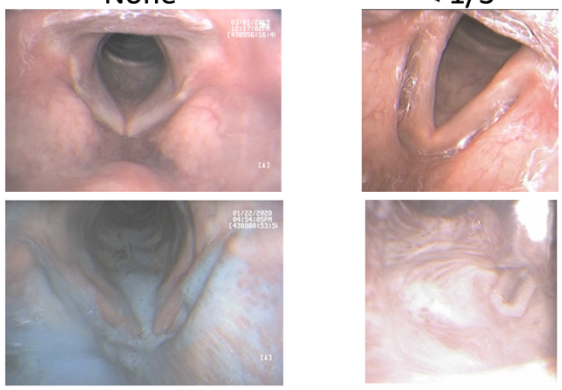
None | < 1/3



1/3 - 2/3 | > 2/3

Vocal Folds (BRACS & VASES)

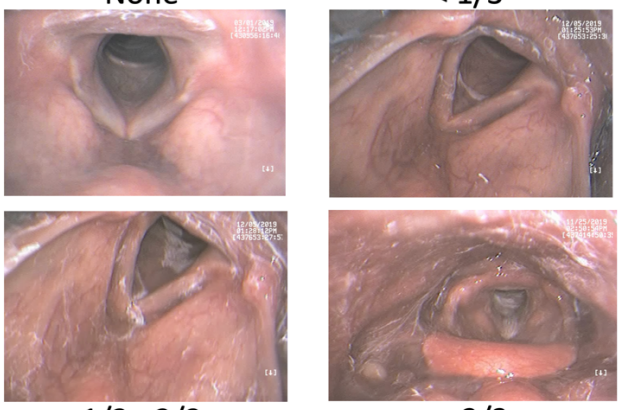
None | < 1/3



1/3 - 2/3 | > 2/3

Subglottis (BRACS & VASES)

None | < 1/3



1/3 - 2/3 | > 2/3