ORIGINAL ARTICLE

Visual Analysis of Swallowing Efficiency and Safety (VASES): A Standardized Approach to Rating Pharyngeal Residue, Penetration, and Aspiration During FEES

James A. Curtis[1] · James C. Borders[1] · Sarah E. Perry[1,2,3,4] · Avery E. Dakin[1] · Zeina N. Seikaly[1] · Michelle S. Troche[1]

Abstract

The primary aim of this study was to describe the development of the Visual Analysis of Swallowing Efficiency and Safety (VASES)—a standardized method to rate pharyngeal residue, penetration, and aspiration during FEES. As a secondary aim, we explored the feasibility of training novices to interpret FEES using VASES. Literature review and consensus panel discussions were used to develop standardized rules for VASES. A training protocol was developed and criterion ratings were established. Twenty-five novice raters completed VASES training and pre-/post-training assessments. Statistical analyses were used to examine pre- to post-training differences in the accuracy, reliability, and time to rate each video clip using VASES. Four sets of VASES rules were developed, including 'what', 'where', 'when', and 'how' to rate FEES. Large, significant post-training improvements in rating accuracy were observed across all seven VASES outcome measures (Cohen's d range 0.74–1.59). Additionally, inter-rater reliability increased for four of the seven outcome measures, and the amount of time to rate each video clip decreased from 2.6 min pre-training to 1.5 min post-training. VASES is a standardized FEES rating method used to enhance the subjective analysis of pharyngeal residue, penetration, and aspiration. It can be feasibly taught to novice raters with a high level of success and may be an effective method to analyze swallowing safety and efficiency in clinical and research practices. Future research is needed to test the validity of VASES by examining its relationship with other validated FEES rating scales.

Keywords Flexible endoscopic evaluation of swallowing (FEES) · Dysphagia · Residue · Aspiration · Visual analysis of swallowing efficiency and safety (VASES)

Introduction

Flexible endoscopic evaluation of swallowing (FEES) is frequently used in clinical and research practice to visualize the pharynx, larynx, and subglottis before, during, and after swallowing [1, 2]. One reason clinicians and researchers use FEES to assess swallowing is to obtain detailed information related to functional swallowing outcomes, including the frequency and severity of impaired swallowing efficiency (pharyngeal residue) and swallowing safety (penetration-aspiration) [2, 3]. This is because current evidence suggests that endoscopic swallowing evaluations possess a sensitivity for assessing pharyngeal residue, penetration, and aspiration that is either comparable to or greater than that of videofluoroscopic swallow studies [4–23].

Ninety to 100% of speech-language pathologists 'usually' or 'always' assess pharyngeal residue, penetration, and aspiration when interpreting instrumental swallowing assessments [24–26]. Despite this, there is a high degree of variability regarding *how* clinicians interpret FEES [13, 27]. In a 2016 survey by Pisegna and Langmore, clinician respondents indicated that the perceived challenges associated with interpreting FEES included: (1) identifying anatomic

✉ James A. Curtis
  jac2406@tc.columbia.edu

1 Laboratory for the Study of Upper Airway Dysfunction, Teachers College, Columbia University, New York City, NY, USA

2 New Zealand Brain Research Institute, Christchurch, New Zealand

3 Department of Medicine, University of Otago, Christchurch, New Zealand

4 School of Psychology, Speech, and Hearing, University of Canterbury, Christchurch, New Zealand

structures; (2) identifying penetration and aspiration; and (3) knowing how to rate residue. In fact, eight methods to rate residue were reported. The majority of respondents (20.3%) reported rating residue based on the amount of residue filling or covering an anatomic structure (an 'anatomically defined residue estimation' method), with a slightly lower number of respondents (18.8%) reporting that they rated residue based on the amount of bolus remaining in the pharynx relative to the estimated amount originally swallowed (a 'bolus clearance estimation' method) [13]. Nearly 17% of the respondents provided vague responses of how residue was rated. This ambiguity and variability in FEES rating methodologies can negatively affect reliability among examiners [28] and makes it difficult to validly compare FEES results across studies. Therefore, there is a significant need to standardize methods for FEES analysis.

Several scales have been developed for, or adapted to, FEES to address these standardization needs [27]. While all of these scales use a categorical rating system, they vary considerably in terms of the outcomes measured, the number of categories used, and the definitions associated with each severity category (Table 1). Additionally, categorical rating methods for FEES [28–38] may not be as reliable or as sensitive as visual analogue scales [39–41] due to their limited ability to document relatively small unit differences [42–44]. In fact, emerging evidence suggests that a visual analogue scale may be a more valid method for rating residue when compared to categorical rating methods [39–41]. Furthermore, there remains ambiguity and inconsistencies related to how anatomic and temporal boundaries are defined within and across these scales, which may limit the reliability and generalizability of findings across studies and clinical practices.

Given the above, the primary aim of this study was to describe the development of the Visual Analysis of Swallowing Efficiency and Safety (VASES)—a standardized approach for rating pharyngeal residue, penetration, and aspiration during FEES. Because standardization of FEES analysis relies on developing a rating method that is feasible to train and implement [45], the secondary aim of this study was to explore the feasibility of training and implementing VASES in a novice group of FEES raters.

Methods

VASES Development

This study was approved by the Teachers College, Columbia University's Institutional Review Board (IRB #: 21-071). The development of VASES included a consensus panel of six ASHA-certified speech-language pathologists who were independent in the performance and interpretation of FEES,

and who had previously published research on topics related to the performance and interpretation of FEES. Three of the panel members were PhD level clinicians with greater than 5 years of treating dysphagia, and three were master's level clinicians with 1–5 years treating dysphagia. All six members obtained dysphagia education and clinical training at differing locations nationally and internationally, including California, Florida, Massachusetts, New York, and New Zealand. VASES development occurred within the context of seven open format discussion meetings. The consensus panel aimed to develop detailed rules and operational definitions for the VASES rating methodology based on: (1) the challenges previously reported for FEES interpretation [13]; (2) the observation that most clinicians use FEES to rate residue, penetration, and aspiration [24–26]; and (3) the presence of inconsistencies and ambiguities within and across scales related to rating scale method (type and description of severity rating levels) and the temporal and anatomic boundaries within which to rate residue, penetration, and aspiration [29–31, 33, 34, 46].

Once these initial rules and operational definitions were established, four of the six panel members re-convened to blindly rate 55 de-identified FEES video clips as a consensus panel using the VASES rules and operational definitions. Each video clip contained a single bolus trial. VASES outcomes were rated within the context open format group discussion using one computer monitor. In a rotating fashion, one of the four panel members would provide initial VASES ratings for all seven outcome measures. Then, the remaining three members of the consensus panel would indicate if they approved the outcome measure ratings or requested a revision to any of the outcome measure ratings. A majority of the panel members (three to four) was required to approve the final rating in order for it to be used as a subsequent criterion reference. The process was repeated for each video clip, with a new member of the consensus panel leading the initial rating. Consensus panel ratings served two purposes. The first purpose was for the consensus panel members to pilot the VASES rating method and to identify areas requiring further refinement. The second purpose was to create criterion references for training proposes.

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 to visualize the pharynx, larynx, and subglottis before, during, and after all swallows. As needed, the endoscope was advanced throughout the

Table 1 Comparisons of FEES Scales

Scale name	Rating outcomes	Rating method	Rating Scale
Yale Pharyngeal Residue Severity Rating Scale (YPRSRS) [29]	Vallecular residue Piriform residue	Anatomically defined residue estimation	5-Point Ordinal Scale 1. None (0% filled) 2. Trace (1–5% filled) 3. Mild (2–25% filled) 4. Moderate (25–50% filled) 5. Severe (>50% filled)
The Pooling Score [28, 33]	Vallecular residue Marginal zone residue Piriform residue Laryngeal vestibule and the vocal folds residue Lower vocal fold residue	Anatomically defined residue estimation	3-Point Ordinal Scale 1. Coating (no definition provided) 2. Minimal (>50% cavity containment) 3. Maximum (>50% cavity containment)
Boston Residue and Clearance Scale (BRACS) [31, 32]	Pharyngeal wall residue Base of tongue residue Vallecular residue Epiglottic tip residue Lateral channel residue Piriform sinus residue Post-cricoid residue Arytenoid residue Aryepiglottic residue Inter-arytenoid residue Laryngeal surface of the epiglottis residue Laryngeal surface of the aryepiglottic fold residue Ventricular folds residue Anterior commissure residue Vocal fold residue Posterior commissure residue	Anatomically defined residue estimation	4-Point Ordinal Scale 1. None/coating (no definition provided) 2. Mild (covering/filling < 1/3) 3. Moderate (covering/filling 1/3–2/3) 4. Severe (covering/filling > 2/3)
Mansoura Fiberoptic Endoscopic Evaluation of Swallowing Residue Rating Scale (MFRRS) [34]	Vallecular residue Piriform residue	Anatomically defined residue estimation	7-Point Ordinal Scale 1. None (no visible residue) 2. Trace (visible coating of the mucosa) 3. Shallow/low pooling 4. Moderate pooling 5. Deep/high pooling 6. Overflowing pooling 7. Penetrating pooling

Table 1 (continued)

Scale name	Rating outcomes	Rating method	Rating Scale
Penetration Aspiration Scale (PAS) [35–37]	Depth of airway invasion Reaction to airway invasion	Anatomically defined residue estimation	8-Point Ordinal Scale 1. Material does not enter the airway 2. Material enters the airway, remains above the vocal folds, and is ejected from the airway 3. Material enters the airway, remains above the vocal folds, and is not ejected from the airway 4. Material enters the airway, contacts the vocal folds, and is ejected from the airway 5. Material enters the airway, contacts the vocal folds, and is not ejected from the airway 6. Material enters the airway, passes below the vocal folds, and is ejected into the larynx or out of the airway 7. Material enters the airway, passes below the vocal folds, and is not ejected out of the trachea despite effort 8. Material enters the airway, passes below the vocal folds, and no effort is made to eject
Dynamic Imaging Grade of Swallowing Toxicity (DIGEST-FEES) [30, 38]	Swallow Efficiency (Pharyngeal residue)	Bolus clearance estimation	4-Point Ordinal Scale (efficiency) 1. Minimal to no residue (< 10%) 2. Less than half residue (10–49%) 3. Majority residue (50–90%) 4. Near complete residue (> 90%)
	Swallowing Safety (Frequency and severity of penetration and aspiration)	Anatomically defined residue estimation	3-Point Ordinal Scale (safety) 1. Trace (faint coating) 2. Neither trace nor gross 3. Gross (> 25% bolus volume)

pharynx and laryngeal vestibule after each swallow to more closely inspect residue patterns throughout the pharynx, laryngeal and subglottic spaces. FEES were completed by, or under the direct supervision of, a speech-language pathologist experienced in the performance and interpretation of FEES. Video clips included thin liquids (IDDSI 0) and puree (IDDSI 4) bolus trials. Thin liquid boluses were impregnated with contrast material to enhance visualization, and included either blue dyed water, green dyed water, white dyed water, or thin liquid barium [47, 48]. Videos included a convenience sample of previously recorded FEES intended to represent a range of rating difficulty as it relates to identifying anatomic and temporal boundaries for FEES interpretation.

Feasibility of VASES Training and Implementation

VASES Training Protocol

Twenty-six novice raters were recruited from a graduate school speech-language pathology department. All raters were master-level students who had one semester of dysphagia coursework with at least one lecture on FEES. The coursework and FEES lecture were from the same instructor for all novice raters. All raters were in the third semester of their training program at the time of the study and had completed one semester of internship training within the university clinic. The internship training did not involve interpretation of FEES or VFSS. Raters were instructed to complete three study phases: (1) Pre-Training Assessment, (2) VASES Training, and (3) Post-Training Assessment. All three phases were completed by the raters using their personal laptop computers in a quiet, private room within their household. Videos used in all three study phases were from the criterion ratings created by the consensus panel VASES development portion of this study.

During the Pre-Training Assessment, novice raters were presented with 25 FEES video clips, with 10 video clips repeated for intra-rater reliability analysis, for a total of 35 video clips. For each video clip, novice raters were instructed to: (1) watch the entire video clip in real-time and then use

slow motion frame-by-frame viewing as needed; (2) refer to the anatomic boundaries reference picture (Fig. 1); (3) rate the most amount of residue seen on anatomic structures from any new bolus material, up until the end of the swallow; and (4) rate the highest PAS score that occurred throughout the entire videoclip. Each novice rater was provided with a copy of the PAS (Table 2) [46]. All novice raters were instructed to complete the ratings within a single sitting.

The VASES Training Phase included five parts completed in sequential order at their own pace over 1 week. Part 1 involved viewing a PowerPoint presentation that outlined the VASES rules and operational definitions (listed below in the Results section). Part 2 involved practicing VASES by viewing and rating five FEES practice video clips (not included in the pre/post assessments). Part 3 involved watching a pre-recorded, 60-min didactic training session between one of the consensus members and a novice rater not involved in this study. Part 4 involved additional VASES practice by viewing and rating another set of five practice video clips. Part 5 involved attending a live, 60-min, 10-person group training session during which the novice raters engaged in a question-and-answer group discussion with one of the

consensus panel members to discuss questions related to the VASES rules.

The Post-Training Assessment involved rating the same set of 35 video clips (re-randomized) that were rated Pre-VASES training. The Post-Training Assessment was completed in one sitting 1 week after VASES Training. Following completion of study participation, the novice raters reported the number of hours required to complete the Pre-VASES Training phase, the VASES Training phase, and the Post-VASES Training phase.

Statistical Analyses

Aim 1: VASES Development Frequency distributions were calculated and used to characterize the VASES criterion ratings of the bolus trials measured by the consensus panel for the pre- and post-training assessments.

Aim 2: Feasibility of VASES Training and Implementation Changes in the accuracy of rating FEES using VASES from pre- to post-training were used as the primary method for examining the feasibility of VASES training. Accuracy was measured by examining the absolute difference in VASES scores between the novice raters' ratings and the consensus panel criterion ratings. Average absolute difference was calculated for each outcome measure. VASES training was considered to be feasible if $\geq 50\%$ of the seven VASES outcome measures demonstrated a post-training increase in VASES accuracy. Wilcoxon signed-rank test was initially used to examine differences in the accuracy of PAS ratings pre- vs. post-training. If data were not normally distributed, then a related-samples sign test was used instead. Paired sample *t* tests were used to examine differences in residue rating scores. Outliers were defined as values greater than 1.5 times the interquartile range above the 75th percentile and below the 25th percentile. If outliers in score differences between the pre- and post-training were detected, or score differences were not normally distributed, then Wilcoxon signed-rank tests were used. Lastly, if data were not symmetric, then a related-samples sign test was used instead of a Wilcoxon signed-rank test. The level of statistical significance was set
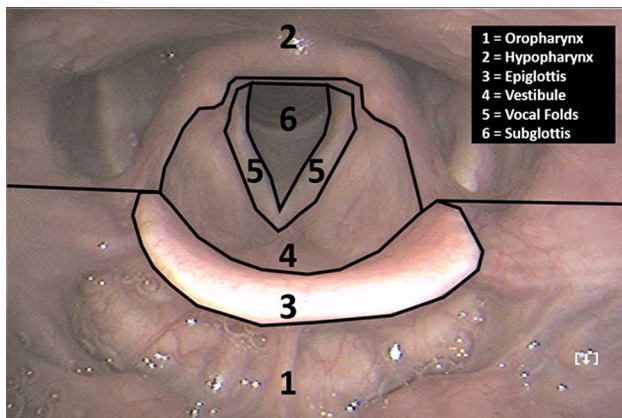

Fig. 1 Picture of the anatomic landmarks provided during pre- and post-training

Table 2 Penetration Aspiration Scale (PAS)

Score	Description
1	Material does not enter the airway
2	Material enters the airway, remains above the vocal folds, and is ejected from the airway
3	Material enters the airway, remains above the vocal folds, and is not ejected from the airway
4	Material enters the airway, contacts the vocal folds, and is ejected from the airway
5	Material enters the airway, contacts the vocal folds, and is not ejected from the airway
6	Material enters the airway, passes below the vocal folds, and is ejected into the larynx or out of the airway
7	Material enters the airway, passes below the vocal folds, and is not ejected out of the trachea despite effort
8	Material enters the airway, passes below the vocal folds, and no effort is made to eject

to a familywise $p < 0.05$. A Holm-Bonferroni adjustment was used to correct for multiple comparisons (i.e., seven comparisons—one for each outcome measure). Cohen's d was used to measure the effect size of training on VASES accuracy for each outcome measure. Effect sizes were interpreted as "small" if $0.2 \leq d < 0.5$, "medium" if $0.5 \leq d < 0.8$, and "large" if $d \geq 0.8$ [49].

Intra- and inter-rater reliability, training completion rate, and the average time to complete VASES training were used as secondary measures to assess the feasibility of VASES training. Forty percent of the videos were randomly selected and repeated for analysis by each novice rater during the pre- and post-training assessments to analyze intra-rater. Inter-rater reliability was calculated for each unique pair of novice raters for each outcome measure. These were then averaged across all pairs of novice raters to characterize average inter-rater reliability for each outcome measure pre- and post-training. Two-way, random effects, intraclass correlation coefficients (ICCs) using absolute agreement were used to examine inter-rater reliability for each of the six residue rating outcome measures. Interpretation of ICC was judged to be 'excellent' if ≥ 0.90, 'good' if between 0.75 and 0.90, 'moderate' if between 0.50 and 0.75, and 'poor' if < 0.50 [50]. Weighted Cohen's Kappa with linear weighting (κ_W) was used to examine intra- and inter-rater reliability for PAS. Interpretation for the κ_W was judged to be 'excellent' if ≥ 0.81, 'good' if between 0.61 and 0.80, 'moderate' if between 0.41 and 0.60, 'fair' if between 0.21 and 0.40, and 'poor' if < 0.20 [50]. Residue reliability ratings that were "good" (ICC ≥ 0.61) [39, 51], and PAS reliability ratings that were "moderate" ($\kappa_W \geq 0.41$) [14, 36, 38, 52–54], were used as cut-offs for training feasibility.

Training completion rate was measured by comparing the number of raters who completed the training relative to the number of raters who started the training. A training completion rate of $\geq 90\%$ was training was selected as a cut-off criterion for training feasibility. The average time to complete VASES training was estimated using self-report from each novice rater. Using the 20–25 h typically required to complete MBSImP training as a referent benchmark [55], an average of 25 h or less to complete VASES training was selected as a criterion for training feasibility.

The time to rate each video clip with VASES was used to measure the feasibility of VASES implementation. This was calculated by dividing the total duration needed to complete each pre- and post-training assessment by the total number of video clips each pre- and post-training phase (i.e., 35 video clips each). The time needed to complete each assessment was tracked and self-reported by each novice rater. Spending an average of 5 min or less to view and rate a single video clip using VASES was selected as the cut-off criteria for what was considered to be feasible for clinical use and implementation.

Results

Aim 1: VASES Development

Rules and operational definitions were created following seven open discussion consensus panel meetings. These rules and operational definitions included four primary areas of analysis, including the *what*, *where*, *how*, and *when* of VASES ratings, as well as additional secondary rules for FEES rating.

'What' to Rate

The first broad area of standardization that emerged from the consensus panel discussions addressed *what* to include in VASES ratings. The PAS was integrated as one of the outcome measures for VASES since it is commonly used in both clinical and research practice to represent impairments in swallowing safety [56, 57]. Residue ratings for six anatomic landmarks were also integrated as outcome measures for VASES. These anatomic landmarks included the oropharynx, hypopharynx, laryngeal surface of the epiglottis, laryngeal vestibule, vocal folds, and subglottis. These structures are frequently seen during FEES [13] and are included in other commonly used FEES rating scales [29–31, 33, 34, 38]. Residue ratings in/on these anatomic landmarks also represent unique impairments in swallowing efficiency (pharyngeal residue) and swallowing safety (penetration and aspiration). Of note, the oropharynx and hypopharynx were used as anatomic boundaries rather than the more typical divisions of the valleculae and piriforms to capture all potential pharyngeal residue not otherwise contained within the valleculae and piriform anatomic landmarks (e.g., on the base of tongue or posterior pharyngeal wall).

'Where' to Rate

The second area of standardization that emerged from the consensus panel discussions involved establishing *where* to delineate exact anatomic landmark boundaries. It was determined that the development of clearly defined anatomic boundaries was necessary for distinguishing oropharyngeal residue from hypopharyngeal residue and for distinguishing between various depths of penetration and aspiration. While many currently available scales provided general descriptions of which anatomic landmarks should be rated for analysis, none of these scales described in detail how to systematically delineate one anatomic landmark from another. Therefore, methods to anatomic boundaries were identified, discussed, and agreed upon by the consensus panel.

Oropharynx-Hypopharynx Anatomic Boundary For the purposes of VASES measurement, the anatomic boundary between the oropharynx and the hypopharynx was first established by identifying the points where the left and right aryepiglottic folds each merge into the laryngeal surface of the epiglottis (i.e., where the medial edge of the aryepiglottic fold is no longer visible). Then, an imaginary horizontal line was drawn across the entirety of the screen connecting these two points (Fig. 2). The space anterior–superior to this line, but not within the laryngeal vestibule, was considered to be the oropharynx, while the space posterior-inferior to this line, but not within the laryngeal vestibule, was considered to be the hypopharynx. An estimation of the imaginary line is extended along the lateral and posterior pharyngeal wall to demark the oro/hypoharyngeal boundary along the pharyngeal wall. The valleculae was defined as the three-dimensional space extending from the tip of the epiglottis (where the lingual and laryngeal surfaces meet) horizontally across to the base of tongue at a perpendicular angle. The piriforms (including the lateral channels) were defined as the three-dimensional space extending from the superior-medial border of the aryepiglottic folds and arytenoids horizontally across lateral and posterior pharyngeal wall at a perpendicular angle.

Laryngeal Surface of the Epiglottis Anatomic Boundary The anatomic boundary between the laryngeal surface of the epiglottis and the laryngeal vestibule was first established by identifying the same two points where the left and right aryepiglottic folds blend into the laryngeal surface of the epiglottis described above. Then, the middle of the trough between the epiglottic petiole and the free end of the epiglottis was identified. Lastly, an imaginary curved line was drawn connecting the points and the trough (Fig. 3). The area anterior–superior to this boundary was considered to be the laryngeal surface of the epiglottis, while the area inferior to this boundary (to the point of the vocal folds) was considered to be the laryngeal vestibule.

Laryngeal Vestibule Anatomic Boundary The anatomic boundary for the laryngeal vestibule was bounded anteriorly superiorly by the laryngeal surface of the epiglottis (described above), and laterally posteriorly by the medial surface of the aryepiglottic folds, arytenoids, and inter-arytenoid tissue (Fig. 4) [52, 58, 59]. Given the ambiguity and inconsistency in rating penetration [2, 46, 60–66], and to maintain generalizability with VFSS, we defined penetration as bolus entering the laryngeal vestibule (PAS 2–3), whereas bolus only on the laryngeal surface of the epiglottis but not within the laryngeal vestibule was not considered penetration (PAS 1).

Vocal Fold Anatomic Boundary The anatomic boundary for the vocal folds included the laryngeal ventricles superiorly, the lateral-most border of the vocal folds, the
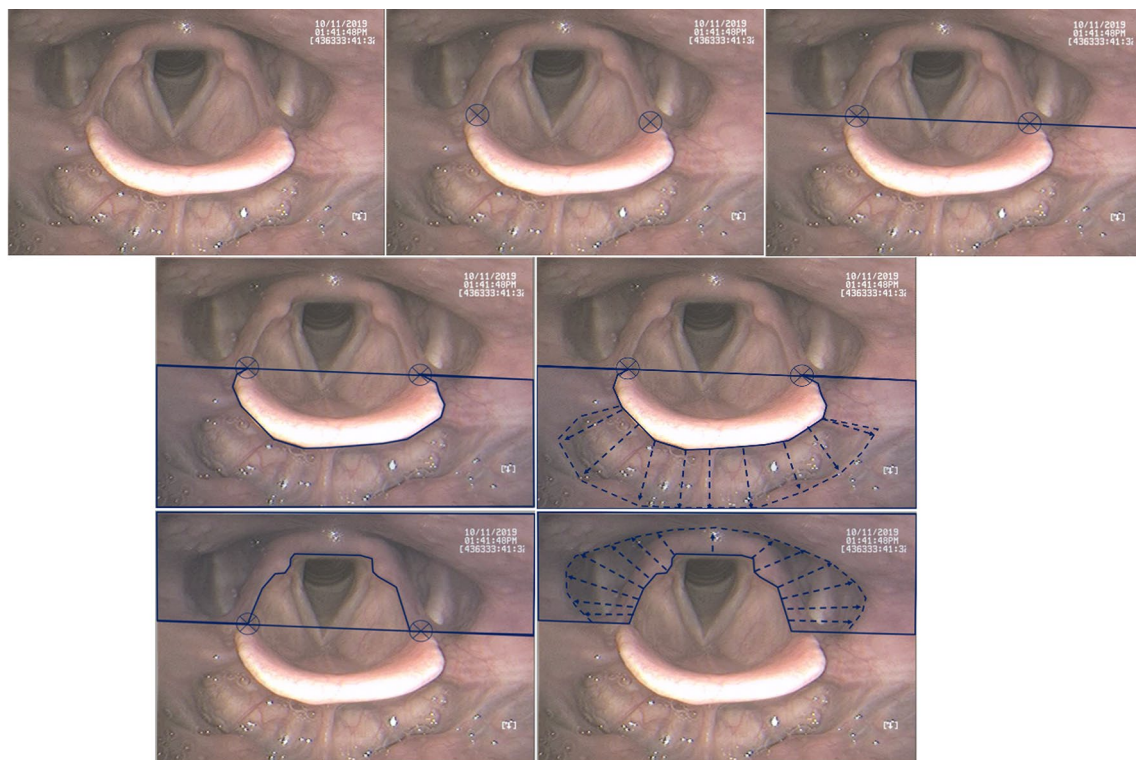

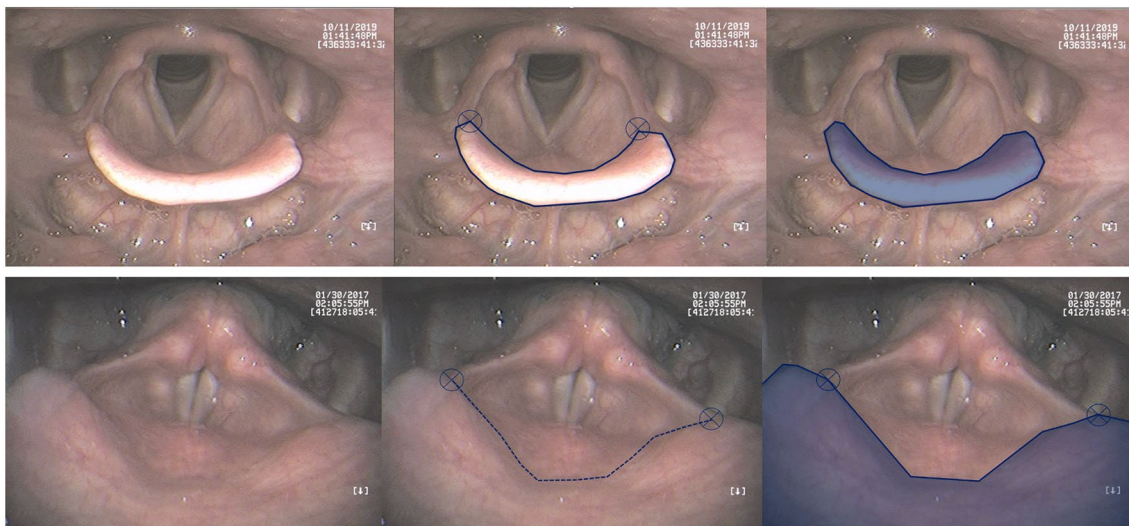**Fig. 2** Anatomic boundary for the oropharynx and hypopharynx

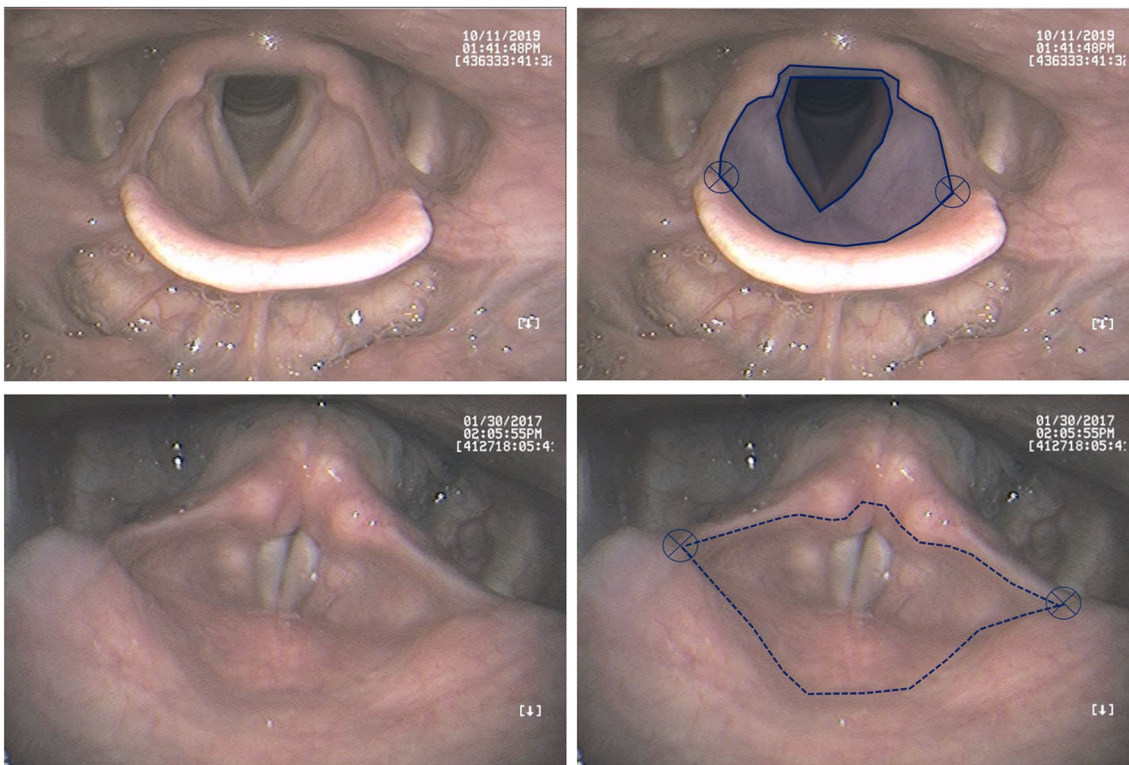**Fig. 3** Anatomic boundary for the laryngeal surface of the epiglottis

Fig. 4 Anatomic boundary for the laryngeal vestibule

cartilaginous portion of the vocal folds posteriorly, and the inferior-most border (i.e., lower lip) of the vocal folds inferiorly. The cartilaginous portion of the vocal folds was determined by extending a line from the superior surface of the vocal fold processes posterior around to the inter-arytenoid tissue (Fig. 5). Bolus contained within the boundaries of the

vocal folds was considered to be penetration to the level of the vocal folds (PAS 4–5) [46]. Aspiration was considered to be present only when bolus crossed the inferior-most border of the vocal folds and into the subglottic space.

Subglottis Anatomic Boundary The anatomic boundary for the subglottis included the subglottic shelf, cricoid
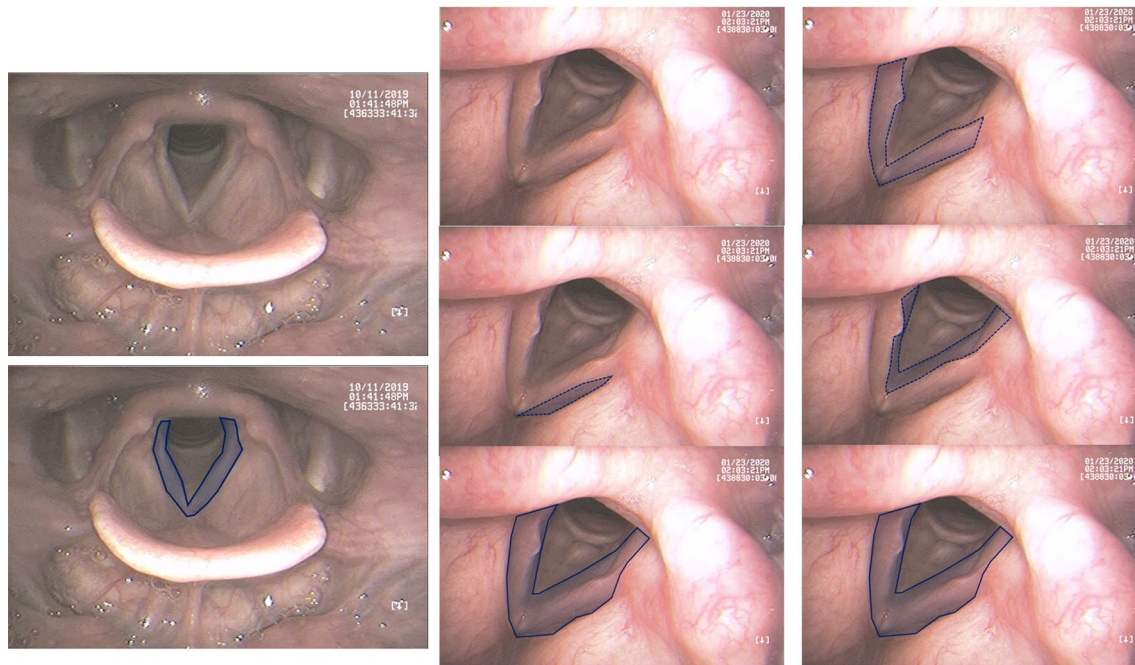
Fig. 5 Anatomic boundary for the vocal folds including the laryngeal ventricles, superior surface of the vocal folds, and medial edge/lower lip of the vocal folds

ring, and trachea (Fig. 6). The subglottic shelf was bounded superiorly and laterally by the inferior-most border of the vocal folds and inferiorly by the superior-most aspect of the cricoid ring. The subglottic shelf extended from the distal (inferior-most) point of the vocal folds down to the proximal (superior-most) border of the cricoid ring. The trachea extended from the distal (inferior-most) border of the cricoid ring down to the carina (if visualized).

'How' to Rate

The third broad area of standardization that emerged from the consensus panel discussions included *how* to complete VASES ratings. Initially, the use of ordinal scales was proposed as a method to rate residue. However, the literature review and open panel discussion revealed three potential drawbacks related to the use of ordinal scales for VASES residue ratings. The first drawback was the lack of an agreed upon number of severity levels. For example, the Pooling score (P-score) contains only three severity levels [33], the Boston Residue and Clearance Scale (BRACS) [31] and the Dynamic Imaging Grade of Swallowing Toxicity for FEES (DIGEST-FEES) [30, 38] contain four severity levels, the Yale Pharyngeal Residue Severity Rating Scale (YPRSRS) contains five severity levels [29], and the Mansoura Fiberoptic Endoscopic Evaluation of Swallowing Residue Rating Scale (MFRRS) [34] contains seven severity levels. The second drawback was the lack of agreement regarding

definitions for residue amount at each severity level. For example, minimal residue for the YPRSRS is equal to 5–25% filled, where minimal residue is < 50% filled for the P-score, < 33% filled for the BRACS, and < 10% remaining for the DIGEST-FEES. The third drawback was the emerging evidence suggesting that residue severity levels differ across bolus consistencies [41], which would necessitate the development of different ordinal rating scales to accommodate ratings of different bolus consistencies.

100-Point Visual Analogue Scale To address the aforementioned drawbacks, the consensus panel adopted a 100-point visual analogue scale to estimate the amount of residue filling or covering each anatomic structure, rather than a general impression how "severe" the residue was perceived to be. The visual analogue scale was used to provide individual residue ratings for each of the six anatomic landmarks. The left-most point of the scale indicated 0% of the anatomic landmark of interest was filled/covered with residue, whereas the right-most point of the scale indicated 100% of an anatomic landmark was filled/covered with residue (Fig. 7). A continuous 100-point scale has the unique advantage of being able to be integrated into pre-existing ordinal rating scales. Research by Pisegna et al. has also identified that visual analogue residue scales yield higher levels of sensitivity and reliability among clinicians when compared to ordinal residue rating scales [39–41]. Lastly, 100-point visual analogue scales have interval statistical properties, which provide greater statistical power when assumptions

are met and can be flexibly adapted to non-parametric alternatives in situations where assumptions are not satisfied.

Anatomically Defined Residue Estimations The consensus panel chose to adopt an anatomically defined residue estimation rating method, as opposed to a bolus clearance estimation rating method, given that the majority of dysphagia clinicians [13] and FEES residue rating scales [29, 31, 33] currently use this rating method (Table 3). Residue ratings for the oropharynx and hypopharynx involved first estimating the amount of residue present anywhere within the oropharyngeal and hypopharyngeal spaces. Then the amount of this residue was expressed as an estimated percentage of how full the valleculae and/or piriforms would be if all of the residue was collected into each of these spaces. Coating of residue on the oropharyngeal or hypopharyngeal mucosa without visible pooling resulted in a residue rating of $\leq 3\%$. Residue ratings for the laryngeal surface of the epiglottis, laryngeal vestibule, and vocal folds involved
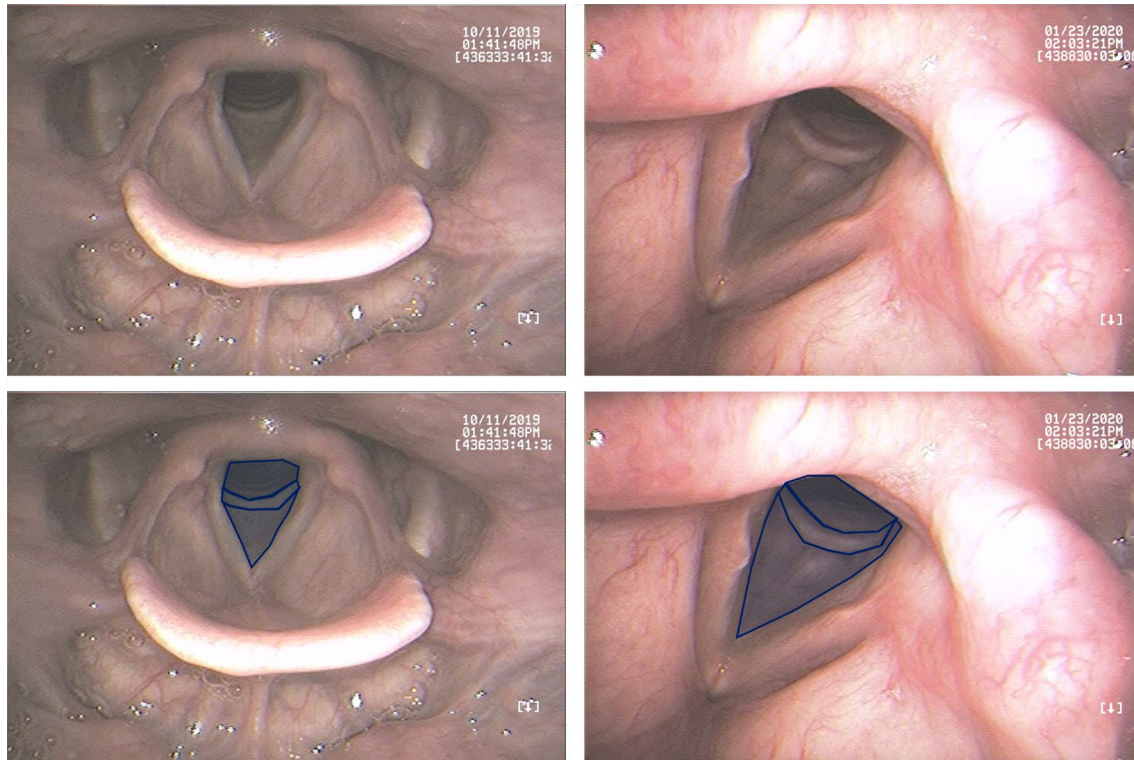

Fig. 6 Anatomic boundary for the subglottis including the subglottic shelf, cricoid ring, and trachea
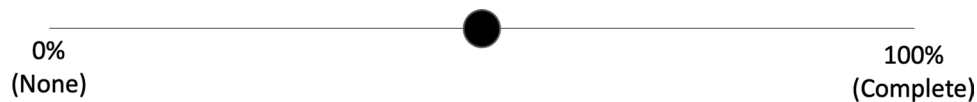

Fig. 7 Example of the visual analogue scale range from 0 (none) to 100% (complete). The central black point (set currently to 50/100) can be moved along the scale

Table 3 Residue rating method		
	Oropharynx	0–100% filled, expressed relative to the vallecular space
	Hypopharynx	0–100% filled, expressed relative to the piriform sinuses and lateral channels
	Epiglottis	0–100% covered, expressed relative to the laryngeal surface of the epiglottis
	Laryngeal vestibule	0–100% covered, expressed relative to the laryngeal vestibule surface area
	Vocal folds	0–100% covered, expressed relative to the vocal folds surface area
	Subglottis	0–100% covered, expressed relative to the subglottic shelf surface area

estimating the amount of mucosa covered with residue and then expressing that as a percentage of the total surface area (visualized and non-visualized) for each anatomic landmark (Fig. 8). Residue ratings for the subglottis were made by estimating the total amount of residue seen covering the subglottic shelf, cricoid ring, and trachea, and then expressing that as a percentage of the subglottic shelf. Only residue that was directly observed, but never inferred, was included in a rating.

'When' to Rate

The fourth broad area of standardization that emerged from the consensus panel discussions involved *when* to rate each of the seven VASES outcome measures. It was hypothesized that this level of standardization was necessary given that FEES allows for continuous visualization of successive bolus trials, often with several minutes of visualization for each bolus trial and the presence and severity of pharyngeal residue, penetration, and aspiration can change as a function of time [60–63]. Because of the influence of time on swallowing efficiency and safety outcome measures, establishing temporal boundaries within which to rate the PAS and residue ratings was warranted.

A literature review revealed four temporal markers within which to rate residue and PAS: before the swallow, during the swallow, after the swallow, and between bolus trials. Taking this literature into account, the following temporal definitions were determined.

- "Before the swallow" began when new foods or liquids entered the mouth and ended at the onset of uninterrupted laryngeal elevation and pharyngeal contraction, leading to the swallowing-related endoscopic whiteout [64, 65].
- "During the swallow" began at the onset of uninterrupted laryngeal elevation or pharyngeal contraction, continued through the period of swallowing-related endoscopic whiteout (when present), and ended when the pharynx and larynx returned to their lowest resting position [52, 63]. In the event of multiple swallows per bolus, 'during the swallow' began at the onset of uninterrupted laryngeal elevation or pharyngeal contraction for the first swallow and ended when the pharynx and larynx returned to their lowest resting positions after the final swallow.
- "After the swallow" began when the pharynx and larynx first returned to their lowest resting positions [52, 63] and ended with the first of any of the following three
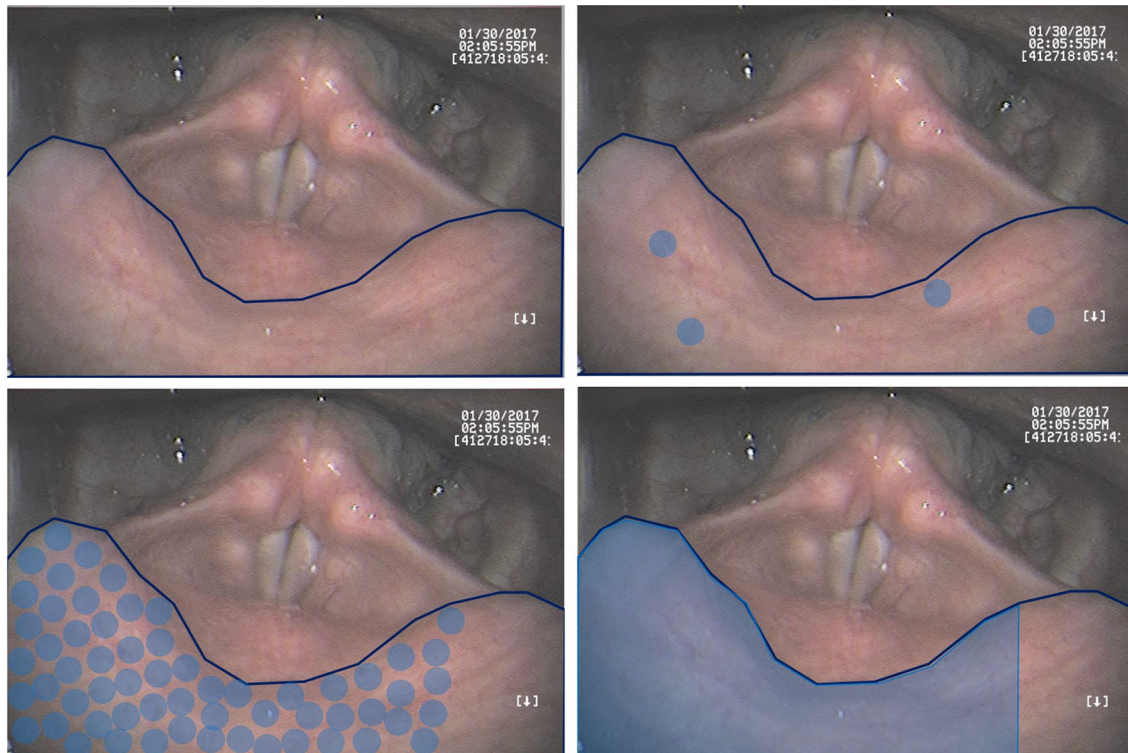

Fig. 8 Illustrated examples of blue residue covering the laryngeal surface of the epiglottis, with residue covering 0% (top left), 3% (top right), 45% (bottom left), and 76% (bottom right)

temporal markers: (1) five seconds of inactivity or rest breathing; (2) patient vocalization (either spontaneous or cued); (3) advancement of the scope towards the laryngeal vestibule.

- "Between bolus trials" included the period of time between "after the swallow" of one bolus trial and "before the swallow" for the subsequent bolus trial.

Residue ratings for the six anatomic landmarks were intended to capture the amount of residue present immediately after the offset of "after the swallow." Therefore, residue was judged before residue flowed from one anatomic landmark to another and before additional cued or spontaneous swallows or coughs were elicited.

Changes in the size and shape of the oropharyngeal and hypopharyngeal spaces can alter the perception of residue severity [66]. For example, the amount of residue may appear greater when the piriforms appear smaller (e.g., when the vocal folds are abducted) compared to when the piriforms appear larger (e.g., when the vocal folds are adducted). Therefore, residue ratings for the oropharynx and hypopharynx were made when the valleculae and piriform sinuses each appeared their largest (e.g., during tidal exhalation or during a sustained phonation task).

Lastly, the single highest PAS score seen across all four temporal markers was used for ratings for each bolus. For example, if vocal fold residue (PAS 5) was initially identified after the swallow, but silent aspiration (PAS 8) was seen between bolus trials from post-swallow residue, then a PAS 8 was given. Conversely, if silent aspiration (PAS 8) was seen immediately after the swallow, leading to a delayed cough between bolus trials (PAS 7), then a PAS 8 was given.

Secondary Rules

Following consensus panel VASES ratings, several secondary rules were developed.

(1) Only judge residue that appears new and is the same color and consistency as the bolus that was presented.

(2) In instances where residue and secretions are mixed together, ignore the secretions and only estimate the amount of residue present.

(3) Raters should explicitly indicate if/when ratings were not made until after coughs or after additional clearing swallows due to incomplete endoscopic visualization, or else decide to systematically exclude such bolus trials from the reported ratings.

(4) Residue that is on, but not crossing, the border of the epiglottis and laryngeal vestibule should be interpreted only as residue that is on the epiglottis and not residue that is in the vestibule.

(5) Residue that is on, but not crossing, the border of the vocal folds and subglottis should be interpreted only as residue that is on the vocal folds and not residue that is in the subglottis.

(6) Residue ratings should ideally be made with a high degree of certainty. If the material being rated could be 'reasonably' interpreted as secretions, shadow, or residue from a previous swallow, then it should not be rated as new residue from the current swallow of interest (Fig. 9).

Consensus Panel Ratings

Consensus panel ratings for the 55 video clips were derived from 15 full-length FEES videos. FEES examinees included ten males and five females with an average age of 69.0 years (SD 8.1). All examinees had a diagnosis of Parkinson's disease, with an average disease duration of 10.2 years (SD 5.7) since symptom onset. Additionally, they had an average SWAL-QOL scaled score of 67.9% (SD 17.9%) [67], indicating moderately impaired swallowing-related quality of life, and a median DIGEST-FEES score of 2 (IQR 2–3; Range 0–4) [30, 38], also indicating moderate dysphagia. Twenty-five video clips were randomly selected from the above FEES and used for assessment of pre- and post-training accuracy. A distribution of the criterion rating for these video clips is outlined in Fig. 10.

Fig. 9 Example of secretions/no vocal fold residue (left), blue residue on the vocal folds (middle), and "bright white" residue on the vocal folds (right)

Aim 2: Feasibility of VASES Training

Pre- vs. Post-Training Accuracy in VASES

Pre-training, novice raters had a tendency to overestimate (rather than underestimate) residue in/on the anatomic landmarks when compared to the criterion ratings. Specifically, 78.1–93.6% of the ratings overestimated residue across the six anatomic landmarks, while only 6.3–21.9% of the ratings underestimated residue across the anatomic landmarks. Additionally, of the incorrect pre-training PAS ratings, 35% had higher ratings and 65% had lower ratings when compared to criterion ratings. Post-training, the types of errors made by the novices were more evenly distributed, with 30.0–53.6% of the ratings having overestimated residue across landmarks, and with 46.3–70.0% of the ratings having underestimated residue across landmarks. Additionally, of the incorrect post-training PAS ratings, 36% had higher ratings and 64% had lower ratings when compared to criterion ratings.

The average absolute error decreased by 48.1–79.6% across the six residue ratings (Fig. 11) while the proportion of incorrect PAS scores decreased by 26.8% (Fig. 12). Results revealed that these were statistically significant improvements (Table 4). Furthermore, medium-to-large effect sizes were observed for the laryngeal vestibule ($d = 0.74$) and subglottic residue ratings ($d = 0.74$), while large effects were observed for the remaining five outcome measures (d range 0.99–1.59).

Secondary Feasibility Measures of VASES Training and Implementation

Reliability Intra-rater reliability improved from "good" to "excellent" for the laryngeal vestibule residue rating, worsened from "good" to "moderate" for the subglottis residue rating, and was unchanged ("good") for the remaining four residue ratings and the PAS. Inter-rater reliability improved from "fair" to "moderate" for PAS, improved from "good" to "excellent" for the hypopharyngeal and laryngeal vestibule ratings, and was unchanged for the oropharynx ("good"),
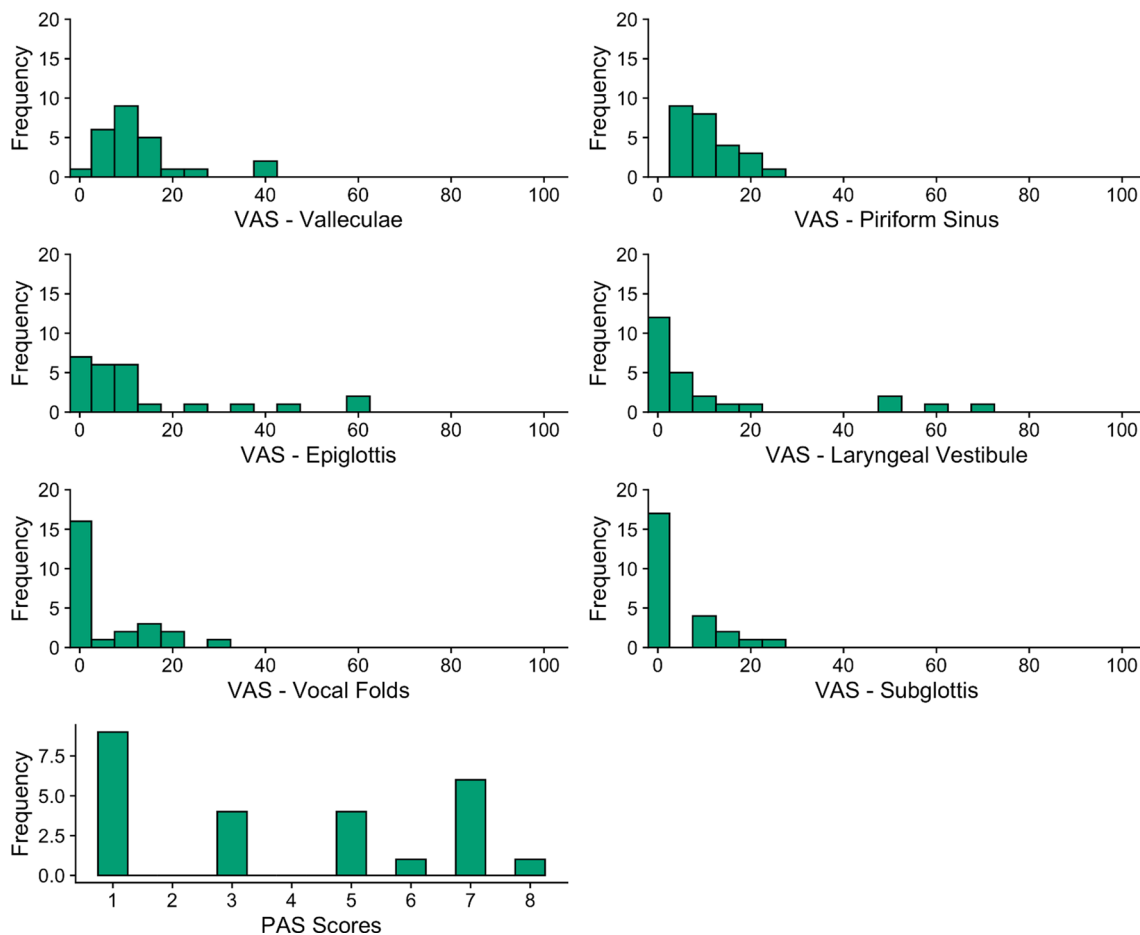

Fig. 10 Frequency distribution of the criterion ratings of the seven VASES outcome measures across the 25 FEES video clips. *Abbreviations*: visual analogue scale ratings (VAS); Penetration Aspiration Scale (PAS)

Fig. 11 Average absolute error in residue ratings across anatomic landmarks for each novice rater pre- and post-training, with lower scores indicating greater accuracy
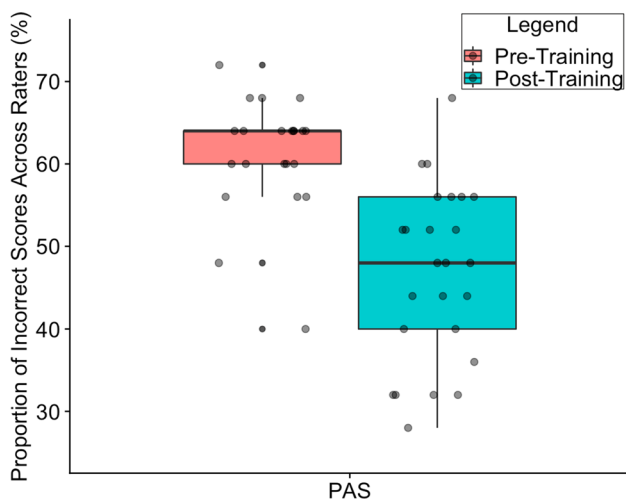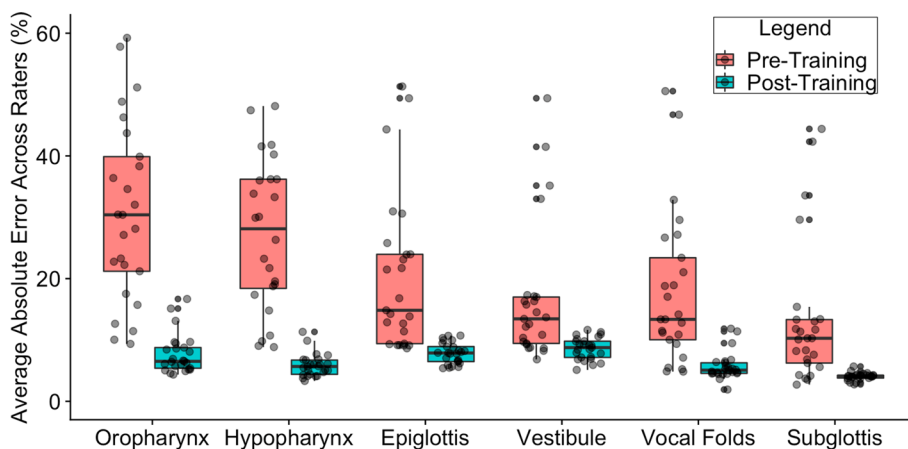

Fig. 12 Proportion of incorrect penetration-aspiration scale (PAS) scores for each novice rater pre- and post-training, with lower proportions indicating greater accuracy

laryngeal vestibule ("excellent"), and subglottis ("good") (Table 5).

Training Completion Rate One of the 26 enrolled novice raters dropped out of the study after beginning the VASES training protocol due to scheduling concerns. The remaining 25 novice raters successfully completed all VASES training, yielding a completion rate of 96.1%.

Time to Complete Training Median number of hours to complete VASES training was 6 h (IQR 4.75–10 h). The minimum time to complete VASES training was 4 h ($n = 6$). Only two raters reported needing more than 10 h to complete VASES training: one of the two raters reported needing 12 h for completion and the other rater reported needing 20 h for completion.

Time to Rate Each Video Clip The videoclips in the Pre- and Post-Training assessments were an average of 1.1 min (SD 0.5). The average self-reported time spent per bolus trial was 3.7 min (SD 1.0) Pre-Training and 2.6 min (SD 1.0) Post-Training. After accounting for the average length of each video, it took an average of 2.6 min to rate each bolus Pre-Training and 1.5 min to rate each bolus Post-Training.

Discussion

While many rating scales have been developed for and adopted to FEES, clearly defined anatomic and temporal boundaries within which to rate pharyngeal residue,

Table 4 Pre- and post-training changes in VASES accuracy

	Pre-training error mean (SD)	Post-training error mean (SD)	Error difference pre- vs post- training mean (SD)	Percent improvement relative to pre-training	Test statistic	P-value	Cohen's d
Oropharynx	30.8 (14.6)	7.7 (3.2)	23.1(14.5)	75.0%	$t = 7.959$	< 0.0005	1.59
Hypopharynx	28.5 (13.7)	5.8 (1.8)	22.7 (14.2)	79.6%	$t = 7.970$	< 0.0005	1.59
Epiglottis	20.2 (12.6)	7.8 (1.7)	12.4 (12.3)	61.3%	$z = -4.346$	< 0.0005	1.00
Vestibule	16.6 (11.0)	8.6 (1.8)	8.0 (10.8)	48.1%	$z = -4.000$	< 0.0005	0.74
Vocal Folds	17.8 (12.2)	5.6 (2.2)	12.1 (12.7)	67.9%	$t = 4.776$	< 0.0005	0.99
Subglottis	13.2 (11.5)	4.0 (0.6)	9.2 (11.6)	69.6%	$z = -2.800$	0.004	0.79
PAS (%)	63.8 (10.2)	46.7 (10.6)	17.1 (13.0)	26.8%	$z = -4.294$	< 0.0005	1.30

Table 5 Pre- and post-training reliability

Outcome	Intra-rater reliability			Inter-rater reliability		
	Average	95% CI	Interpretation	Average	95% CI	Interpretation
Pre-training						
Oropharynx	0.859	0.821–0.889	Good	0.890	0.812–0.994	Good
Hypopharynx	0.877	0.844–0.903	Good	0.896	0.824–0.947	Good
Epiglottis	0.876	0.843–0.902	Good	0.908	0.842–0.953	Excellent
Vestibule	0.911	0.887–0.930	Excellent	0.928	0.876–0.964	Excellent
Vocal folds	0.879	0.847–0.905	Good	0.878	0.796–0.937	Good
Subglottis	0.821	0.773–0.859	Good	0.772	0.630–0.880	Good
PAS	0.605	0.535–0.676	Good	0.294	0.087–0.501	Fair
Post-training						
Oropharynx	0.851	0.811–0.882	Good	0.882	805–0.939	Good
Hypopharynx	0.880	0.848–0.906	Good	0.945	0.907–0.972	Excellent
Epiglottis	0.922	0.902–0.939	Excellent	0.973	0.956–0.986	Excellent
Vestibule	0.943	0.927–0.955	Excellent	0.970	0.950–0.985	Excellent
Vocal folds	0.894	0.866–0.917	Good	0.944	0.906–0.971	Excellent
Subglottis	0.704	0.626–0.767	Moderate	0.849	0.751–0.922	Good
PAS	0.747	0.685–0.809	Good	0.435	0.199–0.671	Moderate

penetration, and aspiration have been lacking. This ambiguity in rating methodology can potentially limit generalizability of research findings and can negatively impact exam reliability. VASES was developed to address these gaps and improve standardization of FEES ratings. Through open panel discussions and a comprehensive literature review, the VASES rating methodology was developed. By describing what, where, when, and how to rate pharyngeal residue, penetration, and aspiration during FEES, clinicians and researchers can begin to measure swallowing efficiency and safety in a more standardized manner.

For a standardized rating method to be widely adopted into clinical and research practices, it must be feasible to train and implement. The primary method for determining if VASES was feasible to train was to examine pre- to post-training differences in the accuracy of VASES ratings. The results from this study demonstrated significant post-training improvements in rating accuracy across all seven VASES outcome measures. However, one limitation to keep in mind was the lack of a "no training" control group. Without a control group, it remains unclear how much of the improvement in rating accuracy was due to training as opposed to general practice of FEES interpretation. However, given that intra-rater reliability was relatively unchanged, but accuracy and inter-rater reliability were improved, current findings suggest that this group of novice raters used a more similar and standardized rating approach post-training compared to pre-training.

From a training feasibility standpoint, we also examined training completion rate, time to complete training, and post-training changes in the accuracy and reliability of ratings. The training completion rate in this group of novice

raters was relatively high (96.1%), and the median time to complete the training was relatively short (6 h), both meeting the cut-off criteria for training feasibility. While 6 h for training may be considered time-consuming by some, it is substantially shorter than the 25 h reported by other popular swallowing training programs [55].

Inter- and intra-rater reliability of all of the VASES outcome measures also met the cut-off for training feasibility, with the exception of PAS which was observed to be "fair" (inter-rater) and "moderate" (intra-rater) post-training. This may have been due in part to the limited number of aspiration events (20%) included in the randomly selected videos for intra-rater reliability assessment. Conversely, this could be an indicator that additional training for PAS may be warranted. It is noteworthy to mention that while PAS reliability was "fair" to "moderate", these values are similar (or greater) to PAS reliability results often reported in the dysphagia literature using similar statistical methods [37, 52–54, 68, 69].

A second important consideration for developing a standardized rating method for FEES is that the rating method must be feasible to implement into clinical practice. This important question is a direction of future research which relies, in part, on identifying facilitators and barriers for implementing standardized rating methods for FEES into clinical practice. While the primary aim of the present study was not designed to examine the feasibility of VASES implementation, the self-reported time to rate a video clip was used to begin to address this clinically relevant question. Reviewing audio–video FEES recordings is an expected and critical component of the evaluation process, is required for FEES billing purposes within the USA [70], and can

significantly influence the reliability and accuracy of FEES interpretation [71]. Findings from the current study revealed that, after training, novice raters took an average of 1.5 min to rate each bolus trial. This met our cut-off threshold of 5 min, suggesting that implementing VASES into clinical practice may be feasible. Furthermore, previous research demonstrates that the time to rate each bolus should decrease over time with continued practice [72]. Therefore, depending on the number of bolus trials and the amount of time logistically allowed for review of the FEES, swallow-by-swallow analysis of FEES using VASES appears largely feasible to implement. Feasibility of implementation may be further increased if analyzing only a select few swallows—an approach often used in clinical practice with other swallowing evaluation protocols (e.g., MBSImP; Dynamic Swallow Study) [64, 73].

It is interesting to note that good to excellent intra- and inter-rater reliability was observed within this cohort of novice raters prior to VASES training. This high level of intra- and inter-rater reliability may be attributed in part to the use of visual analogue scales. This is consistent with the research by Pisegna et al. who identified good-to-excellent reliability when rating pharyngeal residue with visual analogue scales, regardless of experience level [39]. The provision of a standardized anatomic landmark referent image, despite not being provided with the methodology to form standardized anatomic boundaries, may have also contributed to the high level of baseline intra- and inter-rater reliability. Furthermore, reliability appeared relatively unchanged (intra-rater) or improved (inter-rater) after VASES training. This demonstrates that adding the VASES rules and operational definitions does not negatively impact the high level of reliability seen at baseline with use of visual analogue scales and an anatomic boundaries referent image.

Several points of consideration and areas of future research should be considered when interpreting the findings of this study. First and foremost, the present study did not examine the content validity of VASES. Instead, it was intended to describe the development of VASES and to explore its feasibility of use. Given that VASES has now been developed and training has been shown to be feasible, next steps should include examining content validity by comparing VASES to other validated FEES rating scales, including the YPRSRS, BRACS, or DIGEST-FEES. Validation work should also consider exploring the use of VASES in multiple patient populations beyond neurodegenerative disease and VASES sensitivity for tracking changes in swallowing function over time. Second, as part of the VASES training, novice raters engaged in a group Q&A session. By nature, this Q&A session involved some variability in training instruction, potentially limiting the generalizability of training effects. It remains unknown how accuracy could have improved with only providing the written rules (i.e.,

those outlined in this document) or with a standardized set of training documents (see supplemental materials) and not engaging raters in didactic training which can limit potential generalizability. Therefore, another area of future research should consider comparing differences in training methods (e.g., with vs without Q&A sessions). Third, the current study only evaluated the effects of VASES training on novice raters. It is possible that training effects may not generalize to raters with greater FEES experience. Therefore, future research should consider examining if experienced raters can improve the accuracy of VASES ratings following training, and to compare training effects between raters with differing levels of experience. Lastly, VASES is intended to be a standardized framework with which to rate and interpret FEES. It may evolve over time and may be adapted as needed to meet the needs of patient-specific populations. Scales and rating methods which incorporate anatomically defined rules into its protocol (e.g., BRACS, Yale, MBSImP, PAS, VASES) may need to be altered in situations where patients present with altered anatomy (e.g., secondary to cancer or post-operative surgical changes). Adjusting the VASES methodology in these situations is expected. However, if the VASES rating method is adjusted (e.g., by rating left and right oropharynx and hypopharynx separately or by omitting certain outcomes), then these alterations should be specifically reported to maintain transparency, reliability, and standardization.

Conclusions

The Visual Analysis of Swallowing Efficiency and Safety (VASES) is a standardized rating methodology used to enhance the evaluation of pharyngeal residue, penetration, and aspiration on FEES. VASES facilitates good to excellent intra- and inter-rater reliability of FEES analysis in novice raters, can be feasibly taught with high levels of success to novice raters, and may be an effective method to analyze residue, penetration, and aspiration in clinical and research practices. Future research is needed to determine the validity of this method by examining the relationship between VASES ratings with other validated FEES rating scales.

References

1. Langmore SE, Schatz K, Olsen N. Fiberoptic endoscopic examination of swallowing safety: a new procedure. Dysphagia. 1988;2(4):216–9.
2. Langmore SE. History of fiberoptic endoscopic evaluation of swallowing for evaluation and management of pharyngeal

dysphagia: changes over the years. Dysphagia. 2017;32(1):27–38. https://doi.org/10.1007/s00455-016-9775-x.

3. Hiss SG, Postma GN. Fiberoptic endoscopic evaluation of swallowing. Laryngoscope. 2003;113:1386–93.

4. Bevan K, Griffiths MV. Chronic aspiration and laryngeal competence. J Laryngol Otol. 1989;103(2):196–9.

5. Langmore S, Schatz K, Olson N. Endoscopic and videofluoroscopic evaluations of swallowing and aspiration. Ann Otol Rhinol Laryngol. 1991;100(8):678–81. https://doi.org/10.1177/000348949110000815.

6. Curtis JA, Laus J, Yung KC, Courey MS. Static endoscopic evaluation of swallowing: transoral endoscopy during clinical swallow evaluations. Laryngoscope. 2016;126(10):2291–4. https://doi.org/10.1002/lary.25828.

7. Fattori B, Giusti P, Mancini V, et al. Comparison between videofluoroscopy, fiberoptic endoscopy and scintigraphy for diagnosis of oro-pharyngeal dysphagia. Acta Otorhinolaryngol Ital. 2016;36(5):395–402. https://doi.org/10.14639/0392-100X-829.

8. Armstrong ES, Reynolds J, Carroll S, Sturdivant C, Suterwala MS. Comparing videofluoroscopy and endoscopy to assess swallowing in bottle-fed young infants in the neonatal intensive care unit. J Perinatol. 2019;39(9):1249–56. https://doi.org/10.1038/s41372-019-0438-2.

9. Pavithran J, Puthiyottil IV, Kumar M, et al. Exploring the utility of fibreoptic endoscopic evaluation of swallowing in young children—a comparison with videofluoroscopy. Int J Pediatr Otorhinolaryngol. 2020. https://doi.org/10.1016/j.ijporl.2020.110339.

10. Rao N, Brady SL, Chaudhuri G, Donzelli JJ, Wesling MW. Gold-standard? Analysis of the videofluoroscopic and fiberoptic endoscopic swallow examinations. J Appl Res Clin Exp Ther. 2003;3(1):89–96.

11. Kelly AM, Leslie P, Beale T, Payten C, Drinnan MJ. Fibreoptic endoscopic evaluation of swallowing and videofluoroscopy: does examination type influence perception of pharyngeal residue severity? Clin Otolaryngol. 2006;31(5):425–32. https://doi.org/10.1111/j.1749-4486.2006.01292.x.

12. Kelly AM, Drinnan MJ, Leslie P. Assessing penetration and aspiration: how do videofluoroscopy and fiberoptic endoscopic evaluation of swallowing compare? Laryngoscope. 2007;117(10):1723–7. https://doi.org/10.1097/MLG.0b013e318123ee6a.

13. Pisegna JM, Langmore SE. Parameters of instrumental swallowing evaluations: describing a diagnostic dilemma. Dysphagia. 2016;31(3):462–72. https://doi.org/10.1007/s00455-016-9700-3.

14. Scharitzer M, Roesner I, Pokieser P, Weber M, Denk-Linnert DM. Simultaneous radiological and fiberendoscopic evaluation of swallowing ("SIRFES") in patients after surgery of oropharyngeal/laryngeal cancer and postoperative dysphagia. Dysphagia. 2019;34(6):852–61. https://doi.org/10.1007/s00455-019-09979-8.

15. Kamity R, Ferrara L, Dumpa V, Reynolds J, Islam S, Hanna N. Simultaneous videofluoroscopy and endoscopy for dysphagia evaluation in preterm infants—a pilot study. Front Pediatr. 2020. https://doi.org/10.3389/fped.2020.00537.

16. Kaye GM, Zorowitz RD, Baredes S. Role of flexible laryngoscopy in evaluating aspiration. Ann Otol Rhinol Laryngol. 1997;106(8):705–9. https://doi.org/10.1177/000348949710600817.

17. Wu C-H, Hsiao T-Y, Chen J-C, Chang Y-C, Lee S-Y. Evaluation of swallowing safety with fiberoptic endoscope: comparison with videofluoroscopic technique from the departments of otolaryngology. Laryngoscope. 1997;107(3):396–401.

18. Périé S, Laccourreye L, Flahault A, Hazebroucq V, Chaussade S, St Guily JL. Role of videoendoscopy in assessment of pharyngeal function in oropharyngeal dysphagia: comparison with videofluoroscopy and manometry. Laryngoscope. 1998;108(11):1712–6. https://doi.org/10.1097/00005537-199811000-00022.

19. Madden C, Fenton J, Hughes J, Timon C. Comparison between videofluoroscopy and milk-swallow endoscopy in the assessment of swallowing function. Clin Otolaryngol Allied Sci. 2000;25(6):504–6. https://doi.org/10.1046/j.1365-2273.2000.00385.x.

20. Tabaee A, Johnson PE, Gartner CJ, Kalwerisky K, Desloge RB, Stewart MG. Patient-controlled comparison of flexible endoscopic evaluation of swallowing with sensory testing (FEESST) and videofluoroscopy. Laryngoscope. 2006;116(5):821–5. https://doi.org/10.1097/01.mlg.0000214670.40604.45.

21. Singh V, Berry S, Brockbank MJ, Frost RA, Tyler SE, Owens D. Investigation of aspiration: milk nasendoscopy versus videofluoroscopy. Eur Arch Oto-Rhino-Laryngol. 2009;266(4):543–5. https://doi.org/10.1007/s00405-008-0779-5.

22. Da Silva AP, Lubianca NJF, Santoro PP. Comparison between videofluoroscopy and endoscopic evaluation of swallowing for the diagnosis of dysphagia in children. Otolaryngol Head Neck Surg. 2010;143(2):204–9. https://doi.org/10.1016/j.otohns.2010.03.027.

23. Park WY, Lee TH, Ham NS, et al. Adding endoscopist-directed flexible endoscopic evaluation of swallowing to the videofluoroscopic swallowing study increased the detection rates of penetration, aspiration, and pharyngeal residue. Gut Liver. 2015;9(5):623–8. https://doi.org/10.5009/gnl14147.

24. Cimoli M, Oates J, McLaughlin E, Langmore SE. Exploring consistency and variation in fibreoptic endoscopic evaluation of swallowing practice in Australia. Folia Phoniatr Logop. 2019. https://doi.org/10.1159/000503132.

25. Martino R, Pron G, Diamant NE. Oropharyngeal dysphagia: surveying practice patterns of the speech-language pathologist. Dysphagia. 2004;19(3):165–76. https://doi.org/10.1007/s00455-004-0004-7.

26. Swan K, Cordier R, Brown T, Speyer R. Visuoperceptual analysis of the videofluoroscopic study of swallowing: an international Delphi study. Dysphagia. 2020. https://doi.org/10.1007/s00455-020-10174-3.

27. Swan K, Cordier R, Brown T, Speyer R. Psychometric properties of visuoperceptual measures of videofluoroscopic and fibre-endoscopic evaluations of swallowing: a systematic review. Dysphagia. 2018. https://doi.org/10.1007/s00455-018-9918-3.

28. Tohara H, Nakane A, Murata S, et al. Inter- and intra-rater reliability in fibroptic endoscopic evaluation of swallowing. J Oral Rehabil. 2010;37(12):884–91. https://doi.org/10.1111/j.1365-2842.2010.02116.x.

29. Neubauer PD, Rademaker AW, Leder SB. The yale pharyngeal residue severity rating scale: an anatomically defined and image-based tool. Dysphagia. 2015;30(5):521–8. https://doi.org/10.1007/s00455-015-9631-4.

30. Starmer H, Arrese L, Langmore SE, et al. Adaptation of dynamic imaging grade of swallowing toxicity (DIGEST) for fiberoptic endoscopic evaluation of swallowing: DIGEST-FEES. In: *DRS 2020 Annual Meeting*; 2020

31. Kaneoka AS, Langmore SE, Krisciunas GP, et al. The Boston residue and clearance scale: preliminary reliability and validity testing. Folia Phoniatr Logop. 2014;65(6):312–7. https://doi.org/10.1159/000365006.

32. Field K. The Boston Residue and Clearance Scale (BRACS): criterion validity testing. https://hdl.handle.net/2144/12100 (2013). Accessed 1 Nov 2020.

33. Farneti D. Pooling score: an endoscopic model for evaluating severity of dysphagia. Acta Otorhinolaryngol Ital. 2008;28(3):135.

34. Sabry A, Coyle JL, Abou-Elsaad T. Mansoura fiberoptic endoscopic evaluation of swallowing residue rating scale (MFRRS): an anatomically based tool—a preliminary study. Folia Phoniatr Logop. 2020. https://doi.org/10.1159/000512158.

35. Butler SG, Markley L, Sanders B, Stuart A. Reliability of the penetration aspiration scale with flexible endoscopic evaluation of swallowing. Ann Otol Rhinol Laryngol. 2015;124(6):480–3. https://doi.org/10.1177/0003489414566267.

36. Colodny N. Interjudge and intrajudge reliabilities in fiberoptic endoscopic evaluation of swallowing (fees) using the penetration-aspiration scale: a replication study. Dysphagia. 2002;17(4):308–15. https://doi.org/10.1007/s00455-002-0073-4.

37. Nienstedt JC, Müller F, Nießen A, et al. Narrow band imaging enhances the detection rate of penetration and aspiration in FEES. Dysphagia. 2017;32(3):443–8. https://doi.org/10.1007/s00455-017-9784-4.

38. Hutcheson KA, Barrow MP, Barringer DA, et al. Dynamic imaging grade of swallowing toxicity (DIGEST): Scale development and validation. Cancer. 2017;123(1):62–70. https://doi.org/10.1002/cncr.30283.

39. Pisegna JM, Borders JC, Kaneoka A, Coster WJ, Leonard R, Langmore SE. Reliability of untrained and experienced raters on FEES: rating overall residue is a simple task. Dysphagia. 2018;33(5):645–54. https://doi.org/10.1007/s00455-018-9883-x.

40. Pisegna JM, Kaneoka A, Leonard R, Langmore SE. Rethinking residue: determining the perceptual continuum of residue on FEES to enable better measurement. Dysphagia. 2018;33(1):100–8. https://doi.org/10.1007/s00455-017-9838-7.

41. Pisegna JM, Kaneoka A, Coster WJ, Leonard R, Langmore SE. Residue ratings on FEES: trends for clinical application of residue measurement. Dysphagia. 2020. https://doi.org/10.1007/s00455-019-10089-8.

42. Grant S, Aitchison T, Henderson E, et al. A comparison of the reproducibility and the sensitivity to change of visual analogue scales, Borg scales, and Likert scales in normal subjects during submaximal exercise. Chest. 1997;116(5):1208–17.

43. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. J Chronic Dis. 1987;40(12):1129–33.

44. Zraick RI, Kempster GB, Connor NP, et al. Establishing validity of the consensus auditory-perceptual evaluation of voice (CAPE-V). Am J Speech-Lang Pathol. 2011;20(1):14–22. https://doi.org/10.1044/1058-0360(2010/09-0105).

45. Martin-Harris B. Do we have valid and reliable means of quantifying severity of oropharyngeal dysphagia? Moving toward standardization. Perspect Swallow Swallow Disord. 2007;16(1):20–4. https://doi.org/10.1044/sasd16.1.20.

46. Rosenbek JC, Robbins J, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. Dysphagia. 1996;11(2):93–8.

47. Curtis JA, Seikaly ZN, Dakin AE, Troche MS. Detection of aspiration, penetration, and pharyngeal residue during flexible endoscopic evaluation of swallowing (FEES): Comparing the effects of color, coating, and opacity. Dysphagia. 2020. https://doi.org/10.1007/s00455-020-10131-0.

48. Curtis J, Perry S, Troche MS. Detection of airway invasion during flexible endoscopic evaluations of swallowing: comparing barium, blue dye, and green dye. Am J Speech-Lang Pathol. 2019;28(2):515–20. https://doi.org/10.1044/2018_AJSLP-18-0119.

49. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Routledge; 1988.

50. Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.

51. Coffey MM, Tolley N, Howard D, Drinnan M, Hickson M. An investigation of the post-laryngectomy swallow using videofluoroscopy and fiberoptic endoscopic evaluation of swallowing (FEES). Dysphagia. 2018;33(3):369–79. https://doi.org/10.1007/s00455-017-9862-7.

52. Steele CM, Peladeau-Pigeon M, Barbon CAE, et al. Reference values for healthy swallowing across the range from thin to extremely thick liquids. J Speech Lang Hear Res. 2019;62(5):1338–63. https://doi.org/10.1044/2019_JSLHR-S-18-0448.

53. Mancopes R, Peladeau-Pigeon M, Barrett E, et al. Quantitative videofluoroscopic analysis of swallowing physiology and function in individuals with chronic obstructive pulmonary disease. J Speech Lang Hear Res. 2020;63(11):3643–58. https://doi.org/10.1044/2020_JSLHR-20-00154.

54. Waito AA, Plowman EK, Barbon CEA, et al. A cross-sectional, quantitative videofluoroscopic analysis of swallowing physiology and function in individuals with amyotrophic lateral sclerosis. J Speech Lang Hear Res. 2020;63(4):948–62. https://doi.org/10.1044/2020_JSLHR-19-00051.

55. How Long Does the Training Take?. Northern Speech Services. What Is The MBSImP™ Approach? MBSImP | Modified Barium Swallow Impairment Profile. https://www.northernspeech.com/mbsimp/. Accessed 1 Nov 2020.

56. Borders JC, Brates D. Use of the Penetration-Aspiration Scale in dysphagia research: a systematic review. Dysphagia. 2020;35(4):583–97. https://doi.org/10.1007/s00455-019-10064-3.

57. Steele CM, Grace-Martin K. Reflections on clinical and statistical use of the Penetration-Aspiration Scale. Dysphagia. 2017;32:601–16. https://doi.org/10.1007/s00455-017-9809-z.

58. Vose A, Humbert I. "'Hidden in plain sight'": a descriptive review of laryngeal vestibule closure. Dysphagia. 2019;34:281–9. https://doi.org/10.1007/s00455-018-9928-1.

59. Ekberg O. Closure of the laryngeal vestibule during deglutition. Acta Otolaryngol. 1982;93(1–2):123–9.

60. Steele CM, Peladeau-Pigeon M, Barrett E, Wolkin TS. The risk of penetration–aspiration related to residue in the pharynx. Am J Speech-Lang Pathol. 2020;29(3):1608–17. https://doi.org/10.1044/2020_AJSLP-20-00042.

61. 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(4):494–500. https://doi.org/10.1007/s00455-013-9459-8.

62. Lundy DS, Smith C, Colangelo L, et al. Aspiration: cause and implications. Otolaryngol Neck Surg. 1999;120(4):474–8.

63. Pisegna JM, Langmore SE, Meyer TK, Pauloski B. Swallowing patterns in the HNC population: timing of penetration-aspiration events and residue. Otolaryngol Head Neck Surg. 2020. https://doi.org/10.1177/0194599820933883.

64. Martin-Harris B, Brodsky MB, Michel Y, et al. MBS measurement tool for swallow impairment-MBSimp: establishing a standard. Dysphagia. 2008;23(4):392–405. https://doi.org/10.1007/s00455-008-9185-9.

65. Pisegna JM, Langmore SE, Meyer TK, Pauloski B. Swallowing patterns in the HNC population: timing of penetration-aspiration events and residue. Otolaryngol Head Neck Surg (United States). 2020. https://doi.org/10.1177/0194599820933883.

66. Steele CM, Peladeau-Pigeon M, Nagy A, Waito AA. Measurement of pharyngeal residue from lateral view videofluoroscopic images. J Speech Lang Hear Res. 2020;63(5):1404–15. https://doi.org/10.1044/2020_JSLHR-19-00314.

67. McHorney CA, Bricker DE, Kramer AE, et al. The SWAL-QOL outcomes tool for oropharyngeal dysphagia in adults: I. Conceptual foundation and item development. Dysphagia. 2000;15:115–21. https://doi.org/10.1007/s004550010012.

68. Colodny N. Interjudge and intrajudge reliabilities in Fiberoptic Endoscopic Evaluation of Swallowing (Fees®) using the Penetration-Aspiration Scale: a replication study. Dysphagia. 2002;17(4):308–15. https://doi.org/10.1007/s00455-002-0073-0.

69. Langmore SE, Krisciunas GP, Warner H, et al. Abnormalities of aspiration and swallowing function in survivors of acute respiratory failure. Dysphagia. 2020. https://doi.org/10.1007/s00455-020-10199-8.

70. American Speech-Language-Hearing Association. Medicare CPT Coding Rules for Speech-Language Pathology Services. https://www.asha.org/practice/reimbursement/medicare/slp_coding_rules/ (2020). Accessed 24 Dec 2020

71. Hey C, Pluschinski P, Pajunk R, et al. Penetration–aspiration: Is their detection in FEES® reliable without video recording? Dysphagia. 2015;30(4):418–22. https://doi.org/10.1007/s00455-015-9616-3.

72. Nordin NA, Miles A, Allen J. Measuring competency development in objective evaluation of videofluoroscopic swallowing studies. Dysphagia. 2017;32(3):427–36. https://doi.org/10.1007/s00455-016-9776-9.

73. Leonard R, Kendall K. Dysphagia assessment and treatment planning: a team approach. 3rd ed. San Diego: Plural Publishing, Inc.; 2014.

James A. Curtis PhD, CCC-SLP, BCS-S

James C. Borders MS, CCC-SLP

Sarah E. Perry PhD, CCC-SLP

Avery E. Dakin MS, CCC-SLP

Zeina N. Seikaly MS, CCC-SLP

Michelle S. Troche PhD, CCC-SLP