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

James A. Curtis,a James C. Borders,a and Michelle S. Trochea

a Laboratory for the Study of Upper Airway Dysfunction, Department of Biobehavioral Sciences, Teachers College, Columbia University, New York, NY

ARTICLE INFO

Article History:
Received April 20, 2021
Revision received July 29, 2021
Accepted October 25, 2021

Editor-in-Chief: Katherine C. Hustad
Editor: Nicole M. Rogus-Pulia

https://doi.org/10.1044/2021_AJSLP-21-00116

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 analog scale ratings.

Method: Fifty-seven novice raters were prospectively recruited to rate 26 flexible endoscopic evaluations of swallowing (FEES) images (2 times 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 analog scale ratings following verbal numerical ratings. Concurrent validity of using verbal numerical ratings as a potential substitute for visual analog scale ratings was determined by examining the correlation and absolute agreement between both rating methods.

Results: Three thousand five hundred eighty-seven ratings were analyzed. Spearman’s correlation revealed strong correlations between VASES ratings and criterion-referenced ratings across all anatomic landmarks ($\rho = .882-.915$). Lin’s concordance revealed substantial agreement between numerical ratings and visual analog scale ratings ($\rho_c = .986$).

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 analog scales. This suggests that clinicians could provide verbal numerical ratings in lieu of visual analog scale ratings as a potential way to enhance the ease and feasibility of implementing VASES into clinical practice.

Supplemental Material: https://doi.org/10.23641/asha.18737072
clinical and research practices. Preliminary data demonstrate that VASES yields adequate intra- and interrater reliability among 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; Neubauer et al., 2015) and the Boston Residue and Clearance Scale (BRACS; Kaneoka et al., 2014).

The YPRSRS is an anatomically defined image-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), and 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), and 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, Borders, et al., 2018; Pisegna et al., 2020; Pisegna, Kaneoka, et al., 2018; Steele et al., 2020). For example, research by Pisegna and colleagues has found 100-point visual analog scales facilitate greater precision in pharyngeal residue ratings when compared with traditional categorical ratings (Pisegna, Kaneoka, et al., 2018). It is for this reason that visual analog scales were incorporated into the original development of VASES.

### Visual Analog Scales and Verbal Numerical Rating Scales

Visual analog 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 & Skamitzl, 2019; Sussman & Tjaden, 2012; Tjaden et al., 2014; Zraïck et al., 2011). Visual analog scales require measuring the location of a mark on a 100-point line, either digitally or using pen and paper. However, visual analog scales limit the ability to verbally describe impairment to colleagues, which decreases the ease and feasibility of integrating visual analog scales in clinical practice. Given that VASES uses visual analog 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 analog scale ratings. Research comparing visual analog 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 analog scales, then verbal numerical ratings may be a potential substitute for visual analog 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 with 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 that 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 analog scale ratings. We hypothesized that verbal numerical ratings would exhibit substantial agreement with visual analog scales, thus providing evidence that they could be used for rating residue for VASES.

### Method

#### Residue Rating Image Selection

The study was approved by the university’s institutional review board (IRB #: 21–071). Two expert judges (J.C. and J.B.) 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, and 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 Material S1 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) 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, larynx, 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 of these colorants (Curtis et al., 2019, 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, and severe), the percentages associated with each severity rating (0%, 1%–5%, 5%–25%, 25%–50%, and > 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, and 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 prespecified 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

Figure 1. Picture of the anatomic landmarks provided during pre- and posttraining.

Figure 2. Example of the visual analog 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.
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 analog 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 with VASES ratings. Correlations were considered weak if \( \rho < |0.4| \), moderate if \( |0.4| \leq \rho < |0.7| \), strong if \( |0.7| \leq \rho < |1.0| \), and perfect if \( \rho = |1.0| \). A correlation of \( \rho \geq 0.7 \) was set a priori as the cutoff 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 analog 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 analog 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 \( \rho_c > 0.99 \) (Lin, 1989; McBride, 2005; Steichen & Cox, 2002). An agreement of \( \rho_c \geq 0.95 \) was set a priori as the cutoff value for determining if verbal numerical ratings were a valid substitution for visual analog 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.
anatomic landmarks. Specifically, Spearman’s correlation was $\rho = .884$, $p < .0005$ between VASES and YPRSRS for the oropharynx-valleculae (see Figure 3) and $\rho = .893$, $p < .0005$ for the hypopharynx-piriforms (see Figure 4). Additionally, Spearman’s correlation was $\rho = .895$, $p < .0005$ between VASES and BRACS for the epiglottis (see Figure 5), $\rho = .915$, $p < .0005$ for the laryngeal vestibule (see Figure 6), $\rho = .898$, $p < .0005$ for the vocal folds (see Figure 7), and $\rho = .882$, $p < .0005$ for the subglottis (see Figure 8).

### Concurrent Validity: Comparing Verbal Numerical Ratings and Visual Analog Scale Ratings

All 57 novice raters completed these ratings as well, yielding an analysis of 684 concurrent validity ratings. Descriptive statistics of the visual analog scale ratings for each verbal numerical rating are outlined in Table 1. For verbal numerical prompts $\leq 15$ (excluding 0), visual analog scales were an average 1.6 points greater than the verbal numerical prompt. For verbal numerical prompts $\geq 25$ (excluding 100), visual analog 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 analog scale ratings, $\rho_c = .986$ (95% confidence interval [0.984, 0.988]). There was a scale shift of $\omega = 1.034$, a location shift of $\nu = 0.056$, and an accuracy of $\chi_a = .997$ (see Figure 9).

**Figure 3.** Correlation between Visual Analysis of Swallowing Efficiency and Safety (VASES) and Yale Pharyngeal Residue Severity Rating Scale (YPRSRS) for the oropharynx-valleculae.

### 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 judgments of swallowing efficiency and safety as seen during FEES. Previous research has demonstrated that VASES facilitates good-to-excellent reliability among 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’s 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.

**Figure 4.** Correlation between Visual Analysis of Swallowing Efficiency and Safety (VASES) and Yale Pharyngeal Residue Severity Rating Scale (YPRSRS) for the hypopharynx-piriforms.

**Figure 5.** Correlation between Visual Analysis of Swallowing Efficiency and Safety (VASES) and Boston Residue and Clearance Scale (BRACS) for the epiglottis.
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,

Figure 6. Correlation between Visual Analysis of Swallowing Efficiency and Safety (VASES) and Boston Residue and Clearance Scale (BRACS) for the laryngeal vestibule.

Figure 7. Correlation between Visual Analysis of Swallowing Efficiency and Safety (VASES) and Boston Residue and Clearance Scale (BRACS) for the vocal folds.
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 versus nonspecific 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, this study sought to determine if verbal numerical ratings of 0–100 could be used as a valid substitute for the visual analog 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 analog 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 analog 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 this study demonstrated substantial agreement and concurrent validity between verbal numerical ratings and visual analog scales. This suggests

Table 1. Descriptive statistics of visual analog scale ratings across verbal numerical ratings.

<table>
<thead>
<tr>
<th>Verbal numerical rating</th>
<th>Visual analog scale ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>5.0</td>
</tr>
<tr>
<td>5</td>
<td>6.9</td>
</tr>
<tr>
<td>15</td>
<td>16.0</td>
</tr>
<tr>
<td>25</td>
<td>23.4</td>
</tr>
<tr>
<td>33</td>
<td>28.5</td>
</tr>
<tr>
<td>38</td>
<td>31.2</td>
</tr>
<tr>
<td>50</td>
<td>49.7</td>
</tr>
<tr>
<td>66</td>
<td>62.3</td>
</tr>
<tr>
<td>75</td>
<td>71.2</td>
</tr>
<tr>
<td>83</td>
<td>77.2</td>
</tr>
<tr>
<td>100</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Note. CV = coefficient of variation.
that verbal numerical ratings can be confidently used as a valid substitute for visual analog 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 with the criterion-referenced scale within each category because 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 analog scales or 100-point verbal numerical ratings in order to

Figure 9. Agreement between verbal numerical ratings (prompts) and visual analog scale (VAS) ratings, with 57 data points per column. A perfect correlation is represented by the 45° dashed (black) line, whereas the line of best fit for these data is represented by the solid (red) line.
improve the ease and feasibility of implementation into clinical practice.

**Ethical Approval**

All procedures performed were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Approval was obtained from the Institutional Review Board.

**Informed Consent**

Informed consent was obtained from all participants prior to enrollment in this research study.

**References**


