Multicenter Interrater and Intrarater Reliability in the Endoscopic Evaluation of Velopharyngeal Insufficiency

Kathleen C. Y. Sie, MD; Jacqueline R. Starr, PhD; David C. Bloom, MD; Michael Cunningham, MD; Lianne M. de Serres, MD; Amelia F. Drake, MD; Ravindra G. Elluru, MD, PhD; Joseph Haddad Jr, MD; Christopher Hartnick, MD; Carol MacArthur, MD; Henry A. Milczuk, MD; Harlan R. Muntz, MD; Jonathan A. Perkins, DO; Craig Senders, MD; Marshall E. Smith, MD; Travis Tollefson, MD; Jay Paul Willging, MD; Carlton J. Zdanski, MD

Objective: To explore interrater and intrarater reliability ($R_{inter}$ and $R_{intra}$, respectively) of a standardized scale applied to nasoendoscopic assessment of velopharyngeal (VP) function, across multiple centers.

Design: Multicenter blinded $R_{inter}$ and $R_{intra}$ study.

Setting: Eight academic tertiary care centers.

Participants: Sixteen otolaryngologists from 8 centers.

Main Outcome Measures: Raters estimated lateral pharyngeal and palatal movement on nasoendoscopic tapes from 50 different patients. Raters were asked to (1) estimate gap size during phonation and (2) note the presence of the Passavant ridge, a midline palatal notch on the nasal surface of the soft palate, and aberrant pulsations. Primary outcome measures were $R_{inter}$ and $R_{intra}$ coefficients for estimated gap size, lateral wall, and palatal movement; $\kappa$ coefficients for the Passavant ridge, a midline palatal notch on the nasal soft palate, and aberrant pulsations were also calculated.

Results: The $R_{inter}$ coefficients were 0.63 for estimated gap size, 0.41 for lateral wall movement, and 0.43 for palatal movement; corresponding $R_{intra}$ coefficients were 0.86, 0.79, and 0.83, respectively. Interrater $\kappa$ values for qualitative features were 0.10 for the Passavant ridge; 0.48 for a notch on the nasal surface of the soft palate, 0.56 for aberrant pulsations, and 0.39 for estimation of gap size.

Conclusions: In these data, there was good $R_{intra}$ and fair $R_{inter}$ values for qualitative features (ie, the presence of the Passavant ridge, aberrant pulsations, and notch on the nasal surface of the soft palate) is variable.


VELOPHARYNGEAL INSUFFICIENCY (VPI), also known as velopharyngeal (VP) incompetence, inadequacy, or dysfunction, denotes incomplete VP closure during speech production. Clinically, VPI manifests as hypernasal resonance, nasal air emissions, and associated stigmata, including facial grimacing and compensatory misarticulations. The VP complex comprises the levator veli palatini, tensor veli palatini, palatoglossus, palatopharyngeus, muscularis uvula, and the pharyngeal constrictor muscles. Dysfunction of the velopharynx can be caused by anatomic abnormalities, classic cleft palate, and neuromuscular dysfunction.

The primary tool for determining the presence and severity of VPI is the trained ear.\textsuperscript{1,2} Speech pathologists assess resonance, presence of nasal air emissions or turbulence, articulation, and intelligibility. This perceptual assessment of speech is subjective and generally does not predict the degree of VP closure.\textsuperscript{3,4} It is important to establish the relative amount of VP movement because any surgical intervention to improve VP closure carries with it the possibility of compromising the upper airway. Therefore, surgeons have advocated tailoring the surgery to each patient’s needs.\textsuperscript{5-8} Specifically, patients with poor VP closure and a resultant large VP gap require more augmentation than patients with relatively good VP closure re-
In an effort to standardize the reporting of endoscopic and radiographic measures of VP function, a multidisciplinary International Working Group proposed a rating scale in 1988. The Golding-Kushner scale involves separate ratings for the motion of each component of the VP sphincter, that is, right and left lateral pharyngeal and palatal motion, as seen during endoscopic and radiographic evaluations. This scale uses semiquantitative judgments of palatal, lateral pharyngeal, and posterior pharyngeal wall motion during speech relative to rest.

A standardized approach to quantifying, documenting, and recording VPI is critical for reliable communication about diagnosis, clinical care, and research in this field. Because the scale is subject to rater judgment, its usefulness for clinical and research purposes will depend on how variable it is when used by different raters and within raters over time. In the present study, we were interested in estimating intrarater and interrater reliability ($R_{\text{intr}}$ and $R_{\text{inter}}$, respectively) of the Golding-Kushner scale when used by a large number of otolaryngologists who are active in the treatment of VPI across multiple centers in the United States.

**Methods**

Institutional review board exemption was obtained at Children's Hospital and Regional Medical Center (CHRMC), Seattle, Washington. A videotape of nasoendoscopic evaluations of 50 pediatric patients presenting for evaluation of VPI was compiled. The patients were randomly selected from videotapes of all patients with VPI evaluated in the VPI clinic at CHRMC. All nasoendoscopic examinations were performed by 1 of 2 otolaryngologists (J.A.P. and K.C.Y.S.) with a 2.4-mm flexible nasopharyngoscope. The scope was passed through the middle meatus for optimal viewing of the VP during speech, allowing a bird's-eye view of the VP mechanism, generating a 2-dimensional view at the level of attempted VP closure. The endoscopist estimates the degree of lateral pharyngeal and palatal movement during speech. Other characteristics of the VP, such as the presence of a notch on the nasal side of the soft palate, presence of the Passavant ridge, and aberrant pulsations of the internal carotid arteries, can also be visualized when present. These findings may have an impact on patient treatment.

In an effort to standardize the reporting of endoscopic and radiographic measures of VP function, a multidisciplinary International Working Group proposed a rating scale in 1988. The Golding-Kushner scale involves separate ratings for the motion of each component of the VP sphincter, that is, right and left lateral pharyngeal and palatal motion, as seen during endoscopic and radiographic evaluations. This scale uses semiquantitative judgments of palatal, lateral pharyngeal, and posterior pharyngeal wall motion during speech relative to rest.

A standardized approach to quantifying, documenting, and recording VPI is critical for reliable communication about diagnosis, clinical care, and research in this field. Because the scale is subject to rater judgment, its usefulness for clinical and research purposes will depend on how variable it is when used by different raters and within raters over time. In the present study, we were interested in estimating intrarater and interrater reliability ($R_{\text{intr}}$ and $R_{\text{inter}}$, respectively) of the Golding-Kushner scale when used by a large number of otolaryngologists who are active in the treatment of VPI across multiple centers in the United States.

$R_{\text{intr}}$ and $R_{\text{inter}}$ were calculated separately for each individual rater and institution, and then $R_{\text{rater}}$ and $R_{\text{inst}}$ for estimated gap size were calculated separately for segments that were above and below the median duration of nasopharyngoscopic evaluation (80 seconds).

The $R_{\text{rater}}$ and $R_{\text{inst}}$ for assessing qualitative characteristics, including the presence of soft palate notch, aberrant pulsation, Passavant ridge, and categorical gap size, were evaluated by using $\kappa$ coefficients. The $\kappa$ coefficient is a measure of agreement between 2 independent observers that may take on values between 0 and 1. Zero indicates that any agreement observed is completely by chance and 1.0 indicates complete agreement between observers. It is used to describe $R_{\text{rater}}$ and $R_{\text{inst}}$ when data can be classified into nominal categories (eg, “yes” or “no”). In addition to estimating $\kappa$ statistics with bootstrapped 95% confidence intervals, we also calculated, for each of the 50 segments, the number of raters giving positive ratings (of 16 raters). We used only the first of the 2 repeated ratings from each rater for each segment. We plotted the number of segments (of 50) with each possible number of positive ratings (of 16) for each of the 4 qualitative characteristics. All statistical analysis was performed using Stata statistical software (version 9.0; StataCorp, College Station, Texas).

In an effort to standardize the reporting of endoscopic and radiographic measures of VP function, a multidisciplinary International Working Group proposed a rating scale in 1988. The Golding-Kushner scale involves separate ratings for the motion of each component of the VP sphincter, that is, right and left lateral pharyngeal and palatal motion, as seen during endoscopic and radiographic evaluations. This scale uses semiquantitative judgments of palatal, lateral pharyngeal, and posterior pharyngeal wall motion during speech relative to rest.

A standardized approach to quantifying, documenting, and recording VPI is critical for reliable communication about diagnosis, clinical care, and research in this field. Because the scale is subject to rater judgment, its usefulness for clinical and research purposes will depend on how variable it is when used by different raters and within raters over time. In the present study, we were interested in estimating intrarater and interrater reliability ($R_{\text{intr}}$ and $R_{\text{inter}}$, respectively) of the Golding-Kushner scale when used by a large number of otolaryngologists who are active in the treatment of VPI across multiple centers in the United States.

$R_{\text{intr}}$ and $R_{\text{inter}}$ were calculated separately for each individual rater and institution, and then $R_{\text{rater}}$ and $R_{\text{inst}}$ for estimated gap size were calculated separately for segments that were above and below the median duration of nasopharyngoscopic evaluation (80 seconds).

The $R_{\text{rater}}$ and $R_{\text{inst}}$ for assessing qualitative characteristics, including the presence of soft palate notch, aberrant pulsation, Passavant ridge, and categorical gap size, were evaluated by using $\kappa$ coefficients. The $\kappa$ coefficient is a measure of agreement between 2 independent observers that may take on values between 0 and 1. Zero indicates that any agreement observed is completely by chance and 1.0 indicates complete agreement between observers. It is used to describe $R_{\text{rater}}$ and $R_{\text{inst}}$ when data can be classified into nominal categories (eg, “yes” or “no”). In addition to estimating $\kappa$ statistics with bootstrapped 95% confidence intervals, we also calculated, for each of the 50 segments, the number of raters giving positive ratings (of 16 raters). We used only the first of the 2 repeated ratings from each rater for each segment. We plotted the number of segments (of 50) with each possible number of positive ratings (of 16) for each of the 4 qualitative characteristics. All statistical analysis was performed using Stata statistical software (version 9.0; StataCorp, College Station, Texas).

In an effort to standardize the reporting of endoscopic and radiographic measures of VP function, a multidisciplinary International Working Group proposed a rating scale in 1988. The Golding-Kushner scale involves separate ratings for the motion of each component of the VP sphincter, that is, right and left lateral pharyngeal and palatal motion, as seen during endoscopic and radiographic evaluations. This scale uses semiquantitative judgments of palatal, lateral pharyngeal, and posterior pharyngeal wall motion during speech relative to rest.

A standardized approach to quantifying, documenting, and recording VPI is critical for reliable communication about diagnosis, clinical care, and research in this field. Because the scale is subject to rater judgment, its usefulness for clinical and research purposes will depend on how variable it is when used by different raters and within raters over time. In the present study, we were interested in estimating intrarater and interrater reliability ($R_{\text{intr}}$ and $R_{\text{inter}}$, respectively) of the Golding-Kushner scale when used by a large number of otolaryngologists who are active in the treatment of VPI across multiple centers in the United States.

$R_{\text{intr}}$ and $R_{\text{inter}}$ were calculated separately for each individual rater and institution, and then $R_{\text{rater}}$ and $R_{\text{inst}}$ for estimated gap size were calculated separately for segments that were above and below the median duration of nasopharyngoscopic evaluation (80 seconds).

The $R_{\text{rater}}$ and $R_{\text{inst}}$ for assessing qualitative characteristics, including the presence of soft palate notch, aberrant pulsation, Passavant ridge, and categorical gap size, were evaluated by using $\kappa$ coefficients. The $\kappa$ coefficient is a measure of agreement between 2 independent observers that may take on values between 0 and 1. Zero indicates that any agreement observed is completely by chance and 1.0 indicates complete agreement between observers. It is used to describe $R_{\text{rater}}$ and $R_{\text{inst}}$ when data can be classified into nominal categories (eg, “yes” or “no”). In addition to estimating $\kappa$ statistics with bootstrapped 95% confidence intervals, we also calculated, for each of the 50 segments, the number of raters giving positive ratings (of 16 raters). We used only the first of the 2 repeated ratings from each rater for each segment. We plotted the number of segments (of 50) with each possible number of positive ratings (of 16) for each of the 4 qualitative characteristics. All statistical analysis was performed using Stata statistical software (version 9.0; StataCorp, College Station, Texas).
The 16 otolaryngologists had a wide range of experience in evaluating VPI, number of annual VPI evaluations performed, and frequency of use of the Golding-Kushner standardized scale to grade VP function. There was a wide range of level of experience in managing patients with VPI and consistency of use of the Golding-Kushner scale (Table 1).

The 50 segments ranged in duration from 26 seconds to 4 minutes 36 seconds, with a median length of 80 seconds. They included evaluations from 20 girls and 30 boys (mean [SD] age, 9.0 [3.8] years; range, 3.5-17.0 years). This population included 26 patients with cleft palate with or without a cleft lip, who had previously undergone palatoplasty; 15 syndromic patients, including 11 with the diagnosis of velocardiofacial syndrome; 2 patients with VPI following adenotonsillectomy; and 7 patients with idiopathic VPI. The mean (SD) ratings from the 16 otolaryngologists of the VP gap, total lateral wall movement and average palate movement were 23.1%(22.2%), 0.4%(0.3%), and 0.7%(0.3%), respectively (Table 2).

The distribution of estimated gap size ratings across all 16 raters for 10 segments is representative of our find-
ings in all 50 rated segments and demonstrates the variability found in quantifying gap size (Figure). There were segments for which raters differed in estimating gap size by as much as a difference of 80%. The \( R_{\text{inter}} \) and \( R_{\text{intra}} \) for the estimated gap size, total lateral wall movement, and total palatal movement ranged from 0.41 to 0.86 (Table 3). The \( R_{\text{intra}} \) was higher than \( R_{\text{inter}} \) for all 3 measurements. The \( R_{\text{intra}} \) for individual raters varied from 0.54 to 0.99 (Table 4). Neither \( R_{\text{intra}} \) nor \( R_{\text{inter}} \) differed by duration of the segment (whether above or below the median of 80 seconds); \( R_{\text{intra}} \) was no greater for raters by individual or by institutions than among all raters across institutions (data not shown). The \( R_{\text{intra}} \) also did not vary in any systematic way depending on the rater’s historic use of the Golding-Kushner scale (Table 4) or level of experience, and neither did \( R_{\text{inter}} \) (data not shown).

In the absence of an objective, quantifiable measure of VP function, it is important to understand the reliability of clinical tools currently used in the treatment of patients with VPI. Description of VP function is relevant for clinical practice and outcomes assessment. Others have recognized the importance of creating a standardized reporting system. Use of this scale has been variably adopted by otolaryngologists who regularly treat patients with VPI.

In the interest of understanding the current literature on outcomes with VPI surgery, and to design multicenter studies in the future, we used a standardized reporting system to evaluate \( R_{\text{inter}} \) and \( R_{\text{intra}} \) among a large group of otolaryngologists who commonly treat this patient population. We found that the mean \( R_{\text{intra}} \) and \( R_{\text{inter}} \) of gap sizes were approximately 0.86 and 0.63, respectively. The interpretation of these values depends to some extent on the context in which the scale is applied because the “cost” of mismeasurement or disagreement likely differs in different contexts. For example, the use of gap size measurements for determining appropriate surgical treatments might require a higher reliability than if the measurements are made for purely descriptive purposes. Research involving multiple raters might be impractical with an \( R_{\text{inter}} \) of 0.63, which, although considered moderate in some contexts, would necessitate a larger sample size to ensure sufficient statistical power despite potentially large amounts of measurement error.

Other authors have studied the reliability of nasoendoscopic assessment of VP function in subjects with normal VP function or small numbers of subjects. D’Antonio et al conducted a large study of reliability of nasopharyngoscopy prior to recommendations of the International Working Group. Their study was designed to compare the reliability of a group of expert raters, which was their standard clinical practice, with the ratings of novice raters functioning individually. The mean \( \kappa \) statistic was 0.78 for the expert group and 0.51 to 0.50 for the novice group. They also compared the novice ratings with the expert groups, with mean \( \kappa \) scores of 0.3 to 0.4. Although they were able to conclude that experts working together in a group were more consistent than novice raters working individually, we studied a group

**Table 3. Interrater and Intrarater Reliability**

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Reliability Coefficients (95% Confidence Limit Lower Bound)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interrater</td>
</tr>
<tr>
<td>Gap size, %</td>
<td>0.63 (0.54)</td>
</tr>
<tr>
<td>Total lateral wall</td>
<td>0.41 (0.32)</td>
</tr>
<tr>
<td>Mean palate</td>
<td>0.43 (0.31)</td>
</tr>
</tbody>
</table>

For gap percentage, ln-transformed gap percentage, total lateral wall, and mean palate among all raters and all ratings (n=50 segments). See the “Methods” section for a description of the relative rating scale.
of expert raters at different centers, all working independently.

In the present study, all of the participants were asked to use the Golding-Kushner scale to rate the endoscopic segments, regardless of whether they use the scale in their clinical practice. A copy of the article describing the rating scale10 was provided to each rater, but formal training and feedback were not provided. This seems to represent the most typical application of the Golding-Kushner scale because there is no formal or systematic training available regarding its use. All segments were obtained on patients with clinical manifestations of VPI.

This study has several weaknesses. Flexible fiberoptic nasoendoscopic assessments are affected by optical distortion that is difficult to quantify. This distortion renders quantitative measurements impossible. Second, the audio recorded during the examination was included in the videotape, potentially introducing bias based on the degree of VPI. The audio component was included so that raters could correlate VP movement with the production of nonnasal speech sounds. We also felt that inclusion of the audio portion most closely mimicked practical application of the Golding-Kushner scale. Although VPI severity may have influenced the ratings, it would not necessarily have influenced the R_{int} and R_{intra} of the ratings. Another potential source of bias was the format of the samples. The samples were provided to participants in the form of a VHS videotape, and therefore the segments were viewed in the same order during both ratings. We attempted to minimize any recall effects by ensuring that raters waited at least 1 week between making the assessments. With such a large number of segments, it is unlikely that participants would recall their previous rating of any particular segment.

The Golding-Kushner scale itself is a scale of relative movement and therefore is subjective in its application. This scale was chosen because it was created by a multidisciplinary working group. This is the most detailed scale described for assessment of instrumental tests of VP function. It is also appealing because it can be applied to both endoscopic and radiographic tests of the VP.

The strengths of this study include having had an experienced speech therapist present for all the endoscopic examinations to ensure recording of a speech sample tailored to demonstrate optimal function of the VP. With regard to the data set, we developed an Access database that required that all fields be completed by the rater, thereby ensuring a complete data set. Finally, we were able to recruit a large number of raters from geographically diverse centers in the United States.

With these considerations in mind, the R_{int} and R_{intra} for VP gap size were 0.86 and 0.63, respectively. The experience of the rater in terms of number of years of experience and number of patients seen over the previous year did not seem to be associated with the degree of reliability of assessments. Furthermore, the length of the segment also was not associated with reliability. The relatively high R_{intra} of the scale indicates that use of the scale

Table 4. Intrarater Reliability

<table>
<thead>
<tr>
<th>Frequency of Use of Golding-Kushner Scale</th>
<th>Rater No.</th>
<th>Gap, %</th>
<th>Total Lateral Wall Movement</th>
<th>Mean Palatal Movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>10</td>
<td>0.84</td>
<td>0.85</td>
<td>0.67</td>
</tr>
<tr>
<td>Never</td>
<td>1</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>Never</td>
<td>6</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Rarely</td>
<td>2</td>
<td>0.64</td>
<td>0.66</td>
<td>0.54</td>
</tr>
<tr>
<td>Rarely</td>
<td>5</td>
<td>0.91</td>
<td>0.91</td>
<td>0.97</td>
</tr>
<tr>
<td>Occasionally</td>
<td>12</td>
<td>0.60</td>
<td>0.62</td>
<td>0.66</td>
</tr>
<tr>
<td>Occasionally</td>
<td>11</td>
<td>0.77</td>
<td>0.78</td>
<td>0.61</td>
</tr>
<tr>
<td>Usually</td>
<td>9</td>
<td>0.64</td>
<td>0.65</td>
<td>0.72</td>
</tr>
<tr>
<td>Usually</td>
<td>4</td>
<td>0.75</td>
<td>0.76</td>
<td>0.74</td>
</tr>
<tr>
<td>Usually</td>
<td>14</td>
<td>0.92</td>
<td>0.93</td>
<td>0.83</td>
</tr>
<tr>
<td>Always</td>
<td>7</td>
<td>0.77</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>Always</td>
<td>3</td>
<td>0.78</td>
<td>0.79</td>
<td>0.83</td>
</tr>
<tr>
<td>Always</td>
<td>15</td>
<td>0.78</td>
<td>0.79</td>
<td>0.70</td>
</tr>
<tr>
<td>Always</td>
<td>8</td>
<td>0.83</td>
<td>0.84</td>
<td>0.72</td>
</tr>
<tr>
<td>Always</td>
<td>13</td>
<td>0.91</td>
<td>0.92</td>
<td>0.86</td>
</tr>
<tr>
<td>Always</td>
<td>16</td>
<td>0.94</td>
<td>0.94</td>
<td>0.96</td>
</tr>
</tbody>
</table>

\[ ^{a} \text{For assessing gap size (percentage), ln-transformed gap size, total lateral wall, and mean palate, by individual rater (n=50 segments). Note raters are reported by frequency of use of Golding Kushner scale. See the "Methods" section for a description of the relative rating scale.} \]

Table 5. \( \kappa \) Statistics as a Measure of Interrater Reliability

<table>
<thead>
<tr>
<th>Assessment</th>
<th>( \kappa ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passavant ridge</td>
<td>0.10 (0.08-0.13)</td>
</tr>
<tr>
<td>Notch on the nasal side</td>
<td>0.48 (0.45, 0.50)</td>
</tr>
<tr>
<td>Aberrant pulsation</td>
<td>0.56 (0.54-0.59)</td>
</tr>
<tr>
<td>Gap size category</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>0.24</td>
</tr>
<tr>
<td>Small</td>
<td>0.43</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.23</td>
</tr>
<tr>
<td>Large</td>
<td>0.59</td>
</tr>
<tr>
<td>All</td>
<td>0.39 (0.38-0.41)</td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.

\[ ^{a} \text{For assessing Passavant ridge on the nasal side of the soft palate, aberrant pulsation, and gap size category (n=50 segments; 16 raters total).} \]
by a single endoscopist is acceptable. However, the $R_{inter}$ was too low to be used for intercenter comparisons of patient populations at this time. This finding highlights the need to either refine the application of this rating system or to develop a more reliable one.

The variability in the categorical rating of qualitative characteristics is important to address because some authors have suggested using these features to determine candidacy for specific surgical procedures and to define surgical risk. There are inherent limitations to this part of the study owing to our use of $k$ statistics. For example, $k$ statistics are sensitive to the underlying prevalence of a characteristic and thus may not be comparable across studies. Furthermore, a weighted $k$ statistic might have been more appropriate for estimating the reliability of gap size category ratings because weighted $k$ statistics incorporate a penalty scale for disagreements that depends on how great the disagreement (eg, whether disagreements were in adjacent or nonadjacent categories). Thus, these particular $k$ statistics might have underestimated the true reliability of categorical gap size ratings. Nonetheless, the relatively low to moderate reliability observed for categorical characteristics, such as gap size categories and aberrant pulsation, implies a need for refinement of these diagnoses when designing future studies.

A previous study looking at the reliability of naso-endoscopic evaluation has shown that providing feedback during the evaluation or allowing group participation statistically improves reliability. We propose convening a group of otolaryngologists who routinely use this scale to develop a “teaching tape” to demonstrate application of the scale to endoscopic examinations. This could be used as a self-administered tutorial. We will then use a similar design to determine the impact of such training on $R_{inter}$ and $R_{intra}$.

In conclusion, this study demonstrated that the Goldberg-Kushner scale is fairly reliable for describing the findings of a nasoendoscopic evaluation of VP function, making it a useful tool to be used within a center and especially a center in which there is only 1 endoscopist. However, the $R_{inter}$ is currently too low to be used for comparing subjects across centers. It is critical to maximize reliability of a rating scale in the evaluation and reporting of VP function before embarking on multicenter studies.

Submitted for Publication: June 14, 2007; final revision received November 14, 2007; accepted November 25, 2007.

Author Affiliations: Division of Pediatric Otolaryngology, Childhood Communication Center, Children’s Hospital and Regional Medical Center, Seattle, Washington (Drs Sie and Perkins); Departments of Pediatrics and Epidemiology, Children’s Craniofacial Center, Children’s Hospital and Regional Medical Center, University of Washington, Seattle (Dr Starr); Eastern Virginia Medical School and Uniformed Services University Health Sciences, Naval Medical Center Portsmouth, Portsmouth (Dr Bloom); Departments of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston (Dr Cunningham and Hartnick), and New York Medical College, Ardsley (Dr de Serres); Division of Pediatric Otolaryngology, Department of Otolaryngology, University of North Carolina, Chapel Hill, School of Medicine, Chapel Hill (Drs Drake and Zdanski); Division of Pediatric Otolaryngology, Children’s Hospital Medical Center, Cincinnati, Ohio (Dr Elluru and Willging); Departments of Otolaryngology, Children’s Hospital of New York, New York (Dr Haddad), and Oregon Health and Science University, Portland (Drs MacArthur and Milczuk); Division of Otolaryngology, University of Utah, Salt Lake City (Drs Muntz and Smith); and Department of Otolaryngology, University of California, Davis, Sacramento (Drs Senders and Tollefson).

Correspondence: Kathleen C. Y. Sie, MD, Division of Pediatric Otolaryngology, Childhood Communication Center, Children’s Hospital and Regional Medical Center, PO Box 3371/6E-1, Seattle, WA 98105-0371 (kathysie@u.washington.edu).

Author Contributions: All of the authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Sie. Acquisition of data: Sie, Bloom, de Serres, Drake, Elluru, Haddad, Hartnick, MacArthur, Milczuk, Muntz, Perkins, Senders, Smith, Willging, and Zdanski. Analysis and interpretation of data: Sie, Starr, Bloom, Cunningham, Hartnick, Muntz, Senders, and Tollefson. Drafting of the manuscript: Sie, Starr, Bloom, de Serres, Hartnick, and Perkins. Critical revision of the manuscript for important intellectual content: Sie, Starr, Cunningham, Drake, Elluru, Haddad, Hartnick, MacArthur, Milczuk, Muntz, Senders, Smith, Willging, and Zdanski. Statistical analysis: Starr and Hartnick. Administrative, technical, and material support: Sie, Drake, Haddad, MacArthur, and Willging. Study supervision: Sie and Starr.

Financial Disclosure: None reported.

Funding/Support: This study was supported in part by the Murakami Endowment for the Childhood Communication Center, Children’s Hospital and Regional Medical Center, Seattle.

Previous Presentation: This study was presented at the Annual Meeting of the American Society of Pediatric Otolaryngology; May 28, 2005; Las Vegas, Nevada.

Disclaimer: The views expressed in this article are those of the author(s) and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the US government.

REFERENCES


6. Peat BG, Albery EH, Jones K, Pigott RW. Tailoring velopharyngeal surgery; the


