GRBAS and Cape-V Scales: High Reliability and Consensus When Applied at Different Times

Katia Nemr, Marcia Simões-Zenari, Gislaine Ferro Cordeiro, Domingos Tsuji, Allex Itar Ogawa, Maysa Tibério Ubrig, and Márcia Helena Moreira Menezes, São Paulo, Brazil

**Summary:** Objectives. To evaluate whether the overall dysphonia grade, roughness, breathiness, asthenia, and strain (GRBAS) scale, and the Consensus Auditory Perceptual Evaluation—Voice (CAPE-V) scale show the same reliability and consensus when applied to the same vocal sample at different times.

**Methods.** Sixty subjects had their voices recorded according to the tasks proposed in the CAPE-V scale. Vowels /a/ and /i/ were sustained between 3 and 5 seconds. Reproduction of six sentences and spontaneous speech from the request “Tell me about your voice” were analyzed. For the analysis of the GRBAS scale, the sustained vowel and reading tasks of the sentences was used. Auditory-perceptual voice analyses were conducted by three expert speech therapists with more than 5 years of experience and familiar with both the scales.

**Results.** A strong correlation was observed in the intrajudge consensus analysis, both for the GRBAS scale as well as for CAPE-V, with intraclass coefficient values ranging from 0.923 to 0.985. A high degree of correlation between the general GRBAS and CAPE-V grades (coefficient = 0.842) was observed, with similarities in the grades of dysphonia distribution in both scales. The evaluators indicated a mild difficulty in applying the GRBAS scale and low to mild difficulty in applying the CAPE-V scale. The three evaluators agreed when indicating the GRBAS scale as the fastest and the CAPE-V scale as the most sensitive, especially for detecting small changes in voice.

**Conclusions.** The two scales are reliable and are indicated for use in analyzing voice quality.

**Key Words:** Voice disorders–Perceptual ratings–Voice quality–Reliability–GRBAS–CAPE-V–Clinician ratings of voice quality

**INTRODUCTION**

The auditory-perceptual evaluation of voice is one of the most traditional approaches used to analyze voice quality. The evaluation is based on the auditory impression of the evaluator when listening to altered and nonaltered voices, and then it is further compared with physiological findings by a speech therapist. These aspects, added to the patient’s complaints, history of dysphonia, and vocal self-evaluation enable the speech therapist to plan a series of activities aimed at improving the life quality of people with vocal disorders.

Auditory-perceptual evaluation is a highly valued procedure used worldwide and is often regarded as the gold standard for voice disorder documentation. However, the evaluation depends on the experience and training level of the evaluator. The use of evaluation tools, in which predetermined parameters and scales are adopted, is therefore aimed at reducing potential variability and inconsistencies in this analysis.

The GRBAS scale, which assesses overall dysphonia grade, roughness, breathiness, asthenia, and strain, is used worldwide in several fields as a means of vocal evaluation by clinicians and researchers, whereas the Consensus Auditory Perceptual Evaluation—Voice (CAPE-V) also enables the analysis of the same parameters analyzed with the GRBAS scale, except asthenia. Additionally, CAPE-V adopts a visual analog scale and has predetermined vocal tasks and analysis criteria.

GRBAS considers the severity of a vocal disorder along a scale divided in regular intervals, whereas the CAPE-V scale has an asymmetric distribution representing mild, moderate, and severe dysphonia. Both scales are widely used by professionals in the voice field and together they cover all types of voice disorders, regardless of etiology. These include Parkinson’s disease, larynx cancer, granuloma, glottis insufficiency, and Reinke’s edema.

The GRBAS scale is reliable and valid and offers no discomfort or inconvenient to the patient or the therapist. Nevertheless, its sensitivity in detecting vocal alterations is lower than that observed with CAPE-V, probably because GRBAS is an ordinal scale with only three alternatives (mild, moderate, and severe). Indeed, GRBAS has been shown to be less sensitive than CAPE-V to evaluate more subtle differences in vocal quality. On the other hand, CAPE-V represents a shift in the auditory-perceptual evaluation of voice. Its greater sensitivity in detecting small differences in the voice, as compared with GRBAS, has been attributed to its visual analog scale for recording the alteration perception. A slightly improved rater reliability using the CAPE-V to make perceptual judgments of voice quality, in comparison with the GRBAS scale, has also been reported.
CAPE-V has been increasingly used both for research and for clinical practice. Nevertheless, because CAPE-V is a more recent tool than GRBAS, further research is needed to determine the consistency in the assessment produced by these two scales. A previous study that correlated data from both scales, applied at the same time, revealed an agreement among evaluators and a positive correlation between the severity grade of CAPE-V and the overall grade of GRBAS. Karnell et al proposed, however, that further research should be conducted to determine the consistency between the two protocols when applied at different times, as the application of one scale may affect the results obtained when the second scale is applied. The application of the two scales at different times may thus provide the evaluator with the necessary temporal distance to ensure the reliability of the test.

In this study, whether these two voice-related auditory-perceptual evaluation scales show the same reliability and consensus when applied on the same vocal sample at different times was evaluated. The correlation between the two scales was determined considering the severity grade of CAPE-V and the overall grade of GRBAS. The degree of intra- and inter-evaluator agreement was also measured. Additionally, if the Portuguese translation of the CAPE-V scale produces results consistent with those in which this scale was used in its original language was examined, thus determining the validity of its use independently of cultural and language differences.

**METHOD**

The study was approved by the Ethics Committee of the home institution (protocol 0481/08).

An observational cross-sectional design with a sample size of 60 subjects was used (Table 1). Of these, 50 patients had vocal disorders and dysphonia diagnosis, as determined by professionals at the Otorhinolaryngology Clinic, Hospital das Clínicas, Medical School, University of São Paulo, São Paulo, Brazil. Ten volunteers (control group) were also selected among students and patient companions who had no vocal complaints, dysphonia, or laryngeal disorders to enable testing both scales in the absence of vocal alterations.

Patients with different laryngeal diagnoses were selected regardless of gender or age and classified into three dysphonia categories: functional, organofunctional, or organic. Individuals with limited ability to communicate or to be understood were excluded. All participants signed an informed consent.

**Vocal register**

All subjects had their voices recorded according to the tasks proposed in the CAPE-V protocol, using the translation into Portuguese proposed by Behlau. The reading of six sentences and spontaneous speech following the request “Tell me about your voice” were analyzed. For the analysis using the GRBAS scale, the sustained vowel and reading tasks of the sentences was used.

Vocal emissions were recorded in a Sound Forge 6.0 program (Sonic Foundry Inc., Madison, WI), using a Sennheiser brand (model PC-20) headset microphone (Sennheiser, Wedemark, Germany), positioned about 2 in. from the labial commissure of the subject. All recordings were made in an acoustically treated room, with noise below 50 dB.

**Auditory-perceptual voice evaluation**

The auditory-perceptual voice analysis was conducted by three expert speech therapists with more than 5 years of experience and familiar with both the scales.

The final voice sample included the voices of all 60 individuals with 10% random repetition to enable the intra-judge reliability analysis. Voice recordings were stored in two different sequences where the order of the voices in each sample was established randomly. On one sequence, the GRBAS scale was applied (Appendix 1), and on the other, the CAPE-V (Appendix 2) was applied with a 7-day interval from the first evaluation done with GRBAS. The voices were reproduced to the evaluators in a quiet room, through speakers connected to the computer, and were released in three blocks of 20 voices each, with a 10-minute interval between consecutive blocks.

Although the evaluators had experience with both the scales, each evaluator received a review of the guiding parameters and concepts of each scale before the evaluation to ensure high reliability of the evaluations performed by these professionals.

After listening to each voice, each evaluator conducted the analysis individually and no information was shared among evaluators. Each vocal sample could be listened one more time on request. Evaluators were not aware of the patient diagnosis or the fact that individuals with no alteration in voice were included in the sample, as this could compromise their judgment.

In the end of each evaluation, evaluators were asked to indicate the level of difficulty associated with each scale (high, moderate, mild, or absent) and the feasibility of using the scale in the speech therapy clinic (very feasible, feasible, somewhat feasible, or not feasible).

The flowchart below shows all steps taken to evaluate the two scales in this study.
Data analysis

Data analysis was performed using the SPSS program, version 17.0 (Statistical Package for the Social Science; SPSS Inc., Chicago, IL), and a 5% significance level. To determine intra-judge consensus grade (intraclass correlation coefficient calculation), the general vocal alteration grade both from GRBAS and from CAPE-V was used. For interjudge analysis (intraclass correlation coefficient calculation), the following parameters: overall grade, roughness, breathiness, and tension were used from both the scales.

We analyzed the correlation (Spearman-rank correlation test) between the overall GRBAS grade and the CAPE-V severity score as proposed by Karnell et al.¹⁹

RESULTS

Subjects were between 11 and 85 years old, and 72% were women and 28% were men. A strong correlation was observed in the intra-judge consensus analysis, both for the GRBAS scale and for CAPE-V, with intraclass coefficients ranging from 0.923 to 0.985 (Table 2). A high correlation was also observed in the interjudge analysis for all variables (Table 3).

In terms of interscale correspondence, a strong correlation between the overall GRBAS grade and the CAPE-V severity grade was observed (coefficient = 0.842; Figure 1), with similarities in the grades of dysphonia distribution in both scales (Table 4).

The evaluators indicated a mild difficulty in applying the GRBAS scale and low to mild difficulty in applying the CAPE-V scale. Both scales were considered viable or very viable for use. The three evaluators agreed when indicating the GRBAS scale as the fastest and the CAPE-V scale as the most sensitive, especially for detecting small changes in voice.

DISCUSSION

The high levels of intra- and interjudge consensus obtained for the GRBAS and CAPE-V scales emphasize their reproducibility

### Table 2.

<table>
<thead>
<tr>
<th>Judge—Variable</th>
<th>Intraclass Correlation Coefficient</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>J1 G—GRBAS</td>
<td>0.935</td>
<td>0.536</td>
</tr>
<tr>
<td>J1 General Grade—CAPE-V</td>
<td>0.927</td>
<td>0.478</td>
</tr>
<tr>
<td>J2 G—GRBAS</td>
<td>0.923</td>
<td>0.450</td>
</tr>
<tr>
<td>J2 General Grade—CAPE-V</td>
<td>0.985</td>
<td>0.892</td>
</tr>
<tr>
<td>J3 G—GRBAS</td>
<td>0.935</td>
<td>0.536</td>
</tr>
<tr>
<td>J3 General Grade—CAPE-V</td>
<td>0.947</td>
<td>0.619</td>
</tr>
</tbody>
</table>

**Abbreviations:** GRBAS, grade, roughness, breathiness, asthenia, and strain; CAPE-V, Consensus Auditory Perceptual Evaluation—Voice.

### Table 3.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Variable</th>
<th>Intraclass Correlation Coefficient</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower Limit     Upper Limit</td>
</tr>
<tr>
<td>GRBAS</td>
<td>G</td>
<td>0.881</td>
<td>0.817          0.925</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.829</td>
<td>0.737          0.892</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>0.867</td>
<td>0.796          0.916</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>0.793</td>
<td>0.683          0.870</td>
</tr>
<tr>
<td>CAPE V</td>
<td>General grade</td>
<td>0.911</td>
<td>0.864          0.944</td>
</tr>
<tr>
<td></td>
<td>Roughness</td>
<td>0.870</td>
<td>0.800          0.918</td>
</tr>
<tr>
<td></td>
<td>Breathiness</td>
<td>0.897</td>
<td>0.841          0.935</td>
</tr>
<tr>
<td></td>
<td>Tension</td>
<td>0.828</td>
<td>0.735          0.891</td>
</tr>
</tbody>
</table>

**Abbreviations:** GRBAS, grade, roughness, breathiness, asthenia, and strain; CAPE-V, Consensus Auditory Perceptual Evaluation—Voice.
and validity for clinical use, as similarly observed in other publications,\textsuperscript{9–15,18,19,24} and support their continued use. These findings indicate that either scale can be chosen depending on the specific goals of assessment and the time and conditions available for data collection.

Although high in both cases, the interscale correlation found by Karnell et al\textsuperscript{19} was higher than that found in the present study (0.95 vs 0.842). This difference may reflect a reduction in the crossover effect or may be related to the size of the sample in each study (130 in the Karnell et al\textsuperscript{19} study and 60 in the present study). In the present study, however, each scale was used by the same evaluator at a different time, and voices were reproduced randomly, making it unlikely that evaluations were affected by crossover effects.

Although our results support the idea that both scales are suitable for any context where voice-related auditory-perceptual evaluation is relevant,\textsuperscript{15} CAPE-V includes the assessment of additional parameters (pitch, loudness, two further parameters, resonance classification, and display of additional features), thus being suitable for more complete evaluations and consequently a broader understanding of vocal patterns. The presence of predefined procedures for data collection and evaluation also facilitates its implementation and documentation of samples, allowing greater exchange of information between experts. This is especially relevant in view of the fact that the CAPE-V scale has not yet been widely used in many countries, so our results encourage its use. Owing to its great applicability, the use of the GRBAS scale should not be, however, interrupted. Indeed, all evaluators were unanimous in stating that both scales meet the needs of voice auditory-perceptual evaluation and that they can be selected according to specific requirements. The GRBAS scale provides more promptitude and objectivity, with focus on the glottic level, regardless of sample type, whereas the CAPE-V scale considers more detail and analytical parameters, with predefined voice sample collection and evaluation.

In this study, the Portuguese-translated version of CAPE-V\textsuperscript{20} was used, similarly to some previous studies.\textsuperscript{21,22} The fact that our findings are similar to those in which the scale was used in its original language helps validate the Portuguese-translated version of the CAPE-V scale and supports the idea that the scale can be used in different populations, regardless of cultural and language differences.

Professional experience depends on the quality of practice and time dedicated to it. Therefore, exposure to auditory-perceptual analysis should be reinforced in speech therapist qualification. In a study carried out with undergraduates, auditory training involving the GRBAS scale improved the students’ initial skills and prepared them better to perform evaluations.\textsuperscript{25} Considering the potential for the increased use of the CAPE-V scale as based on the present findings, it remains to be determined whether these latter results are also valid for CAPE-V. It is also necessary to develop a curriculum for graduate students, which includes theory and hands-on training focusing on both scales.

### TABLE 4.
Distribution of Findings According to the Classification of GRBAS Scale General Grade in Relation to the Score Regarding Severity Grade of CAPE-V Protocol

<table>
<thead>
<tr>
<th>General Grade CAPE-V</th>
<th>N</th>
<th>G0</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1–9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10–19</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>20–29</td>
<td>9</td>
<td>0</td>
<td>7</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>30–39</td>
<td>9</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>40–49</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>50–59</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>60–69</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>70–79</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>80–89</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>90–99</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>0</td>
<td>15</td>
<td>30</td>
<td>15</td>
</tr>
</tbody>
</table>

\textit{Abbreviations:} GRBAS, grade, roughness, breathiness, asthenia, and strain; CAPE-V, Consensus Auditory Perceptual Evaluation—Voice. Bold values indicate numeric distribution equivalence of general classification of GRBAS grade scale and the score regarding severity grade of CAPE-V protocol.
Finally, it should be noted that the high correlation levels found in this study refer to those parameters common to both scales (namely GRBAS overall grade and CAPE-V severity grade). Still, voice auditory-perceptual evaluation is part of the multidimensional evaluation process, so possible associations among voice auditory-perceptual, physiological, and acoustic parameters may also provide valuable information to researchers in this field.

CONCLUSION

There is high reliability and consensus between the GRBAS and CAPE-V scales, supporting the idea that they can be both used in any situation in which voice-related auditory-perceptual evaluation is relevant. Both scales meet the needs of voice auditory-perceptual assessment and can be selected depending on specific practical, clinical, or research requirements.

REFERENCES

APPENDIX 1

CAPE-V Form

Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)

Name: __________________________ Date: __________________________

The following parameters of voice quality will be rated upon completion of the following tasks:
1. Sustained vowels, /a/ and /i/for 3-5 seconds duration each.
2. Sentence production:
   a. The blue spot is on the key again.
   b. How hard did he hit him?
   c. We were away a year ago.
   d. We eat eggs every Easter.
   e. Mymama makes lemon muffins.
   f. Peter will keep at the peak.
3. Spontaneous speech in response to: “Tell me about your voice problem,” or “Tell me how your voice is functioning.”

Legend: C = Consistent  I= Inconsistent  
MI = Mildly Deviant  MO= Moderately Deviant  
SE= Severely Deviant

Overall Severity
   MI  MO  SE  
   C  I  /100

Roughness
   MI  MO  SE  
   C  I  /100

Breathiness
   MI  MO  SE  
   C  I  /100

Strain
   MI  MO  SE  
   C  I  /100

Pitch
   (Indicate the nature of the abnormality): __________________________
   MI  MO  SE  
   C  I  /100

Loudness
   (Indicate the nature of the abnormality): __________________________
   MI  MO  SE  
   C  I  /100

COMMENTS ABOUT RESONANCE:
   NORMAL  OTHER (Provide description): __________________________

ADDITIONAL FEATURES (for example, diplophonia, fry, falsetto, asthenia, aphonia, pitch instability, tremor, wet/gurgly, or other relevant items):

Clinician: __________________________

APPENDIX 2

GRBAS Scale

GRBAS Scale

Protocol number: ________ Voice number: ________ Date: ____ / ____ / ______

G____ R____ B____ A____ S____