Development and Rationale for the Consensus Auditory-Perceptual Evaluation of Voice—Revised (CAPE-Vr)

[#a]Gail B. Kempster, [†]Kathleen F. Nagle, and [‡,§,¶]Nancy Pearl Solomon, #*Chicago, IL*, †*South Orange, NJ*, ‡§*Bethesda, and* ¶*College Park, MD*

Summary: Rationale. The Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) has been in circulation for more than 20 years. Over the course of time, issues have arisen that have had an impact on the intended administration and interpretation of this common clinical tool.

Purpose. Based on published literature, clinical experience, recent survey data, and practical considerations, and while maintaining the original purpose of the instrument, the authors developed a revised protocol, new rating form, and updated instructions for the CAPE-V, now called the CAPE-V—Revised (CAPE-Vr).

Summary of Modifications. Revisions to the CAPE-V include the following: removal of textual labels indicating regions of severity under each visual analog scale on the rating form, instead displaying terms indicating the direction of the lines; modification of several of the stimuli; revised rating options for pitch, loudness, and resonance, and an added category for nasality; added space to describe inconsistencies according to task; modified options for vocal instabilities and other features; and added space for comments about overall impression. The form also includes sections for documenting recording and rating conditions. Updated instructions are provided to clarify the CAPE-Vr protocol and correspond closely to the rating form.

Conclusion. The CAPE-Vr is constructed to avoid common errors and problems identified from previous use of the original CAPE-V. This paper provides a rationale for each modification to the original CAPE-V, an updated form, and an example of a completed form. The CAPE-Vr is intended as a clear and useful assessment tool for documenting the auditory-perceptual evaluation of voice.

Key Words: CAPE-V—CAPE-Vr—Voice quality—Assessment—Voice evaluation.

INTRODUCTION

The Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)[1] was developed under the auspices of American Speech-Language-Hearing Association (ASHA) Special Interest Division (SID) 3: Voice and Voice Disorders (now Special Interest Group (SIG) 3: Voice and Upper Airway Disorders) by a group of speech-language pathologists, voice scientists, and psychoacousticians. The authors of the CAPE-V were interested in creating a brief but thorough and reliable tool for clinically evaluating voice in a standard way[2,3] based on discussions and research presented at the conference. At the time, the 4-point GRBAS scale[4] was used most often in the international community to assess and document perceived voice quality according to five dimensions: grade, roughness, breathiness, asthenia, and strain.[5–7] The ordinal GRBAS scale, developed in Japan, was considered inadequate for research as it lacked specific instructions and provided only three discrete ordinal indicators of severity (1 = mild, 2 = moderate, 3 = severe) for abnormality within each dimension. The desired outcome of

the CAPE-V was to introduce a way for clinicians and researchers to consistently document perceived voice qualities and associated factors in order to promote improved communication, understanding, and replicability among professionals involved with the assessment of voice disorders.

BACKGROUND

The CAPE-V was drafted immediately following an international consensus conference devoted to the topic of the auditory perception of voice quality, held at the University of Pittsburgh on June 10-11, 2002. It utilized a psychometrically valid and sensitive set of visual-analog rating scales. The protocol intentionally sampled voice using three types of tasks (sustained vowels, reading sentences, conversational speech) to capture a range of production styles. The instructions detailed how to record the tasks, including mouth-to-microphone distance, sampling rate, and ambient noise level, and how to document results on a comprehensive single-page form.

The CAPE-V was introduced to the ASHA community on the SID 3 website in 2002 and published in the American Journal of Speech-Language Pathology (AJSLP) by Kempster and colleagues in 2009.[1] Since its publication, the CAPE-V has become a commonly used protocol for evaluating voice quality in the U.S. and has been translated and adapted for use around the world.[8] In fact, an expert panel of laryngologists, mostly from the United States, named the 2009 publication as one of the 21 most influential papers in laryngology since the year 2000.[9]

The CAPE-V is often considered the "gold standard" protocol for auditory-perceptual evaluation of voice, which

is arguably the most useful component of voice examinations.[10] As such, it is used as a comparison to other metrics. Early validation studies compared it to the GRBAS[2,3] and to the self-assessment questionnaires *Voice-Related Quality of Life* and *Iowa Patient's Voice Index.*[2] These studies supported the CAPE-V as a valid tool for rating voice quality and revealed that self-assessments evaluate a different construct. The CAPE-V is taught in graduate speech-language pathology programs and is published in numerous American textbooks. The CAPE-V has become nearly ubiquitous with auditory-perceptual evaluation of voice in the U.S.

The ASHA Publications Office receives frequent requests to translate the CAPE-V into different languages. One major aspect of adapting the CAPE-V internationally is the development of linguistically and culturally appropriate sentence stimuli. We are aware of validated versions that have been published in Hindi,[11] Mandarin,[12] Kannada,[13] Italian,[14] French,[15] Spanish,[16] bilingual Catalan/Spanish,[17] Japanese,[18] Tamil,[19] Malay,[20] two Turkish versions,[21,22] and three Portuguese versions—two in European Portuguese[23,24] and one in Brazilian Portuguese.[25] It is likely that even more translations and adaptations of the CAPE-V protocol and form exist and continue to be created.

With its extensive use over two decades, a variety of issues have emerged that have complicated the use of the CAPE-V for both researchers and clinicians. This article has two primary purposes: to review the literature regarding the current use of the CAPE-V, thereby elucidating elements of continuing concern, and to present a revision of the CAPE-V along with updated instructions for its administration.

ISSUES WITH THE CAPE-V

We have identified three general topics of concern that affect the administration of the CAPE-V: 1) discrepancies between the original CAPE-V rating form initially posted on the SIG 3 website (2002, re-dated 2009) and the form published in the appendix of the peer-reviewed publication describing its development;[1] 2) inconsistent and incorrect implementation of the protocol; and 3) outdated instructions based on efficiency and current technology. Each of these issues detract from the primary purpose of the CAPE-V, which was to provide clinicians and researchers with a *standard and systematic* assessment of the auditory perception of voice quality. The following sections address each of these issues through the lens of understanding the nature and impact of the problems so that solutions could be developed.

CAPE-V Forms: Descriptions and Discrepancies

The form used with the original CAPE-V protocol consisted of eight scales, one for overall quality, three for specific descriptors of voice quality (roughness, breathiness, and strain), one each for pitch and loudness, and two unspecified to allow raters to list additional parameters as appropriate. Each parameter was rated along an undifferentiated 100-mm horizontal line. In the pure sense, this is known as a visual analog scale (VAS). With the CAPE-V, however, textual markers were included beneath each line to indicate approximate locations for ordinal categories of severity (MI = mild; MO = moderate; SE = severe), leading to a hybrid of a VAS and nonlinear ordinal scale.[26,27] The scales for pitch and loudness included additional space to indicate the nature of the abnormality. Further, users were prompted to select "consistent" (C) or "intermittent" (I) for each parameter. In addition, users were instructed to rate different voice tasks separately if they noticed differences across tasks. Finally, there are sections on the CAPE-V form to note whether resonance was normal or not with a space to elaborate and room to add additional features (examples provided are "diplophonia, fry, falsetto, asthenia, aphonia, pitch instability, tremor, wet/gurgly, or other relevant terms").

The instructions included with the CAPE-V protocol indicated that the clinician may place tick marks at any point along the VAS. Measurement in millimeters from the left end of the line was to be reported as a proportion of the 100-mm line. Textual labels placed under the lines were meant to indicate general regions of severity. Protocol instructions suggested using both numeric ratings and categorical descriptors of severity when reporting results.

Soon after the peer-reviewed publication of the CAPE-V protocol,[1] SID 3 members noted discrepancies between the original form posted on the ASHA website in 2002, which we hereafter call *Form 1* (Figure 1), and the form published in the appendix of the article,[1] referred to here as *Form 2* (Figure 2). After a series of email communications regarding this issue between two of us (NPS and GK) in 2010, Kempster consulted with the original consensus group and responded that "the form as it is published in AJSLP is very close in its dimensions to the original form drafted by the authors in June of 2002 … after considerable reflection, we have decided that the form published in AJSLP should be viewed since publication as the official form. It represents the intent of the authors of the paper. Forms available prior to the publication of the 2009 article may be considered 'beta forms'."

As is evident in Figures 1 and 2, the primary difference between the forms was the placement of the ordinal text markers for severity—MI, MO, and SE—beneath the VAS lines. On Form 1,[28] these were centered at about 10, 35, and 72 mm, respectively, on each of the eight scales, resulting in a non-equidistant placement that was intended to represent the common exponential relationship between perceptual ratings and acoustic correlates of sound. This resulted in little room to indicate normal-to-mild dysphonia and a relatively large area of the scale dedicated to severely deviant voice quality.[26] Form 2[1] centered the labels at about 10, 50, and 90 mm such that MO was in the middle of the line and MI and SE were about 10 mm from the left and right ends of the line, respectively. According to Kempster and colleagues,[1] psychometric expert L. Marks recommended that anchors on a continuous graphical (visual

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. My mama 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 = Intermittent
> MI = Mildly Deviant
> MO = Moderately Deviant
> SE = Severely Deviant

SCORE

Overall Severity _____ C   I   ____/100
                  MI   MO   SE

Roughness   _____ C   I   ____/100
                  MI   MO   SE

Breathiness   _____ C   I   ____/100
                  MI   MO   SE

Strain   _____ C   I   ____/100
                  MI   MO   SE

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

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

_____   _____ C   I   ____/100
                  MI   MO   SE

_____   _____ C   I   ____/100
                  MI   MO   SE

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

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

Clinician:_____

FIGURE 1. Original CAPE-V form (Form 1) released in November, 2002.[28,45] Reprinted with permission from ASHA.

analog) scale should not be placed at the scale's endpoints but rather should leave room at either end to avoid endpoint effects. The consensus group also noted that the "qualitative terms [were] positioned in a nonequidistant fashion, based on Marks's recommendations" (p. 127).[1] The form included in the publication (Form 2) followed Marks's original recommendation (leaving space at the ends of the scales) but *not* the group's interpretation of his advice (nonequidistant positioning). In an interim paper describing the development of the CAPE-V, Kempster[29] explained that "The unequal positioning of these terms is deliberate and based upon findings in psycho-acoustic and

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.   d. We eat eggs every Easter.
    b. How hard did he hit him?   e. My mama makes lemon muffins.
    c. We were away a year ago.   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 = Intermittent
>   MI = Mildly Deviant   MO = Moderately Deviant   SE = Severely Deviant
> **Although the PDF scale is accurate, printer configurations vary. Verify that your paper copy has accurate 100-mm lines before reproducing this form.**

Overall Severity _____   C   I   ___/100
                  MI   MO   SE

Roughness   _____   C   I   ___/100
                  MI   MO   SE

Breathiness   _____   C   I   ___/100
                  MI   MO   SE

Strain   _____   C   I   ___/100
                  MI   MO   SE

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

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

_____   _____   C   I   ___/100
                  MI   MO   SE

_____   _____   C   I   ___/100
                  MI   MO   SE

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

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

                                          Clinician:_____

FIGURE 2. CAPE-V form (Form 2) published in May, 2009.[1] Reprinted with permission from ASHA.

psychometric research." (p.12) This rationale was re-iterated by Kempster et al[1] despite the change in the labels' placements in the article's appendix. The article also included a statement in the acknowledgments that "the form and protocol included in this article as Appendices B and C have been modified slightly from the initial version" (p. 128) without specifying what had changed. [Other changes to the form included reformatting the legend, a warning to users to verify that the lines are 100 mm in length, and permission to photocopy the form for clinical purposes.] The CAPE-V's creators' decision that the forms were similar enough to proceed with the use of Form 2 resulted in both forms being in circulation. This inevitably led to confusion and complications for reporting results.

It is important to note that the validation studies involving the CAPE-V[2,3] and a popular clinical-simulation training site[30] used Form 1. For many years, Form 1 was reprinted in textbooks and taught in graduate school classes. Over time, several textbooks have replaced it with Form 2 or some iteration thereof. Form 2 appears in the 10th edition of the popular textbook by Boone and colleagues.[31] Both Ferrand[32] and Stemple et al[33] show modified versions of the CAPE-V with symmetrical labels (per Form 2) in their latest editions.

The existence of two different rating forms has also had an impact on research efforts. Form 1 has been used in research studies to examine the effects of training and experience on voice assessment,[26,34] and to describe a wide variety of voice disorders, from pediatric[35] to postsurgical[36] to adult-neurogenic dysphonia.[37,38] Many studies cite Kempster et al[1] as the source for the CAPE-V, but unless it is specified, it is impossible to confirm which form was actually used. Consequently, Nagle[39] noted that both versions of the form are available and that "care must be taken to use the same version if repeated ratings are obtained" (p. 49).

Researchers have grappled with the label-location issue in various ways. For example, Awan and Lawson[34] placed the severity labels as shown in Form 1, ie, shifted towards the left side of the line; yet they instructed their listeners thusly: "Since, severity exists on a continuum ranging from normal found on the left end of the scale to severe on the right side of the scale, you are urged to use the entire line when making your judgments." (p 351). Other authors who used Form 1 also provided specific instructions to use the entire line.[26] Recent studies examining perceptual ratings of voice based on the CAPE-V have eliminated the labels entirely, instead opting to present raters with unmarked lines.[40–42] Kania et al[43] examined regions of severity on the unmarked VAS scales by also having expert listeners rate severity with the ordinal GRBAS scale for all 296 voices. Using receiver-operating characteristic analysis, they determined ranges for overall severity, breathiness, and roughness but were unable to reliably determine severity ranges for strain. The cut-off values separating the severity levels (normal, mild, moderate, and severe) generally matched the location of the MI, MO, and SE severity markers from Form 1. Notably, however, the original intent of the markers was to indicate general regions of severity, not cutoff values between them.[1]

This discrepancy between the forms led Nagle et al[44] to investigate whether the location of the textual markers affected ratings of dysphonic voices. Inexperienced listeners rated the overall severity of samples of dysphonic speech using 100-mm lines with three variations regarding severity labels: labels under the line placed nonlinearly as in Form 1, labels under the line placed symmetrically as in Form 2, and a traditional VAS with no labels under the line but instead with the words "normal" and "extremely severe" placed to the left and right of the line, respectively.[44] Although mean ratings were generally comparable across the

three varieties of scales, there was a sex difference such that listeners who used Form 1 judged female voices as significantly less severe than those who used either Form 2 or the traditional VAS. This indicated that, at least in some cases, ratings were driven leftward, ie, towards normal, when the "MO" and "SE" labels reflected the exponential distribution of perceived severity. Rather than accommodating a larger potential range of "severe" ratings, the nonlinear placement of textual markers effectively reduced the portion of the scale used by raters. These findings supported the impression described by Solomon and colleagues[27] that "the data are artificially compressed by imposing the nonlinear auditory-perceptual system on a nonlinear visual perception system" (p. e12). Furthermore, the reliance on the labels may even be more pronounced for experienced clinicians as compared to novice listeners.[26]

Meanwhile, the question of which form to use has complicated international use of the adapted CAPE-V forms. As reviewed by Mahalingam and colleagues,[8] the European Portuguese version[24] and Hindi version used Form 1,[11] whereas the Kannada,[13] Turkish,[21,22] and Portuguese[23] versions used Form 2. The Mandarin version[12] changed the scales from VAS to equal-appearing intervals, and the bilingual Catalan/Spanish version removed the labels entirely.[17] Calaf and Garcia-Quintana used the 4-point GRBAS as well as both versions of the CAPE-V rating form[17] to demonstrate that the labels on Form 1 were good indicators of cut-off values between severity categories, which was not the original intent, and that the "moderate" label, but not the "mild" or "severe" labels, on Form 2 represented the general regions of corresponding impairment. Given the difficulties interpreting the regions of severity, these authors opted to remove severity labels as well as endpoint anchors in their version of the CAPE-V.

Overall, despite emphases in the protocol instructions from 2002 and 2009 that "A key issue is that the regions indicate *gradations* in severity, rather than discrete points,"[28,45] there was the potential, and indeed the strong temptation, for raters to use the textual labels situated below the VAS lines as anchors. With two existing, highly circulated, differing versions of the CAPE-V scales, the probability of unreliable, invalid, and nonspecific results has been high. Interestingly, only one of 17 SLPs surveyed regarding their use of the CAPE-V in 2015 reported awareness that two separate versions of the rating form existed.[46] If users were unaware of this issue, they would not be alert to the need to report which form was used and may have selected a different version for re-evaluation. As can be understood through the history recounted here, use of a clearly understood standard rating form—an essential element of any assessment tool—was not established for the CAPE-V. This continues to result in confusion when reporting and interpreting CAPE-V results.

CAPE-V Protocol: Implementation Problems
The authors of the CAPE-V published guidelines regarding the proper and expected administration of the protocol and

completion of the rating form.[1,28] Since then, it has become clear that most clinical users do not faithfully follow those guidelines.[46–48] Likely reasons involve time constraints by clinicians in busy clinical settings, the perceived value of the information obtained, and unfamiliarity with the recommended protocol.

The vast majority of surveyed SLPs reported completing CAPE-V administration and scoring in under 10 minutes, with no more than 5 minutes allotted to soliciting the stimuli and 5 minutes for rating and scoring.[46,48] However, clinicians are often expected to complete a full voice assessment in under an hour, including taking a case history, conducting perceptual, acoustic, videostroboscopic, and possibly aerodynamic testing,[49,50] discussing findings, providing recommendations, and documenting results. It is no surprise, therefore, to learn that many users do not follow the protocol as published, instead opting to eliminate certain tasks in an apparent effort to shorten administration time.

Tasks and scoring elements of the CAPE-V may also be skipped or altered because clinicians do not perceive their inherent value to the diagnostic process. Although the auditory-perceptual evaluation of voice quality is indisputably central to all voice assessments, the tasks required for this purpose are less clear cut.[50] Rationales for the tasks and scoring for the CAPE-V are provided by Kempster et al (2009); for example, to maintain validity and reliability, ratings should be based on audio recordings of the CAPE-V tasks rather than other components of the voice examination. However, users frequently select the tasks and rating methods that seem most relevant for them and their patients at any given time.[48]

We have been examining issues of administration and scoring procedures over the past decade. Lodhavia and Kempster[46] surveyed 17 voice-specialized SLPs about their typical administration and scoring of the CAPE-V in 2015. In 2019, Nagle[47] directly observed 20 voice-experienced SLPs while they scored audio samples of dysphonic voices using Form 2 of the CAPE-V according to their usual practice. Most recently, we surveyed voice-focused SLPs at the 2023 Fall Voice Meeting in Washington, DC[48]; up to 59 individuals responded to queries about their typical strategies for the administration and scoring of the CAPE-V. From these studies, we can describe the actual usage of the CAPE-V and examine possible trends over time. To our knowledge, no other research groups have addressed these issues systematically, although in a cross-cultural systematic review of studies involving translations of the CAPE-V, Mahalingam and colleagues[8] noted differences in the CAPE-V rating procedures across 10 studies that reported the use of this tool.

CAPE-V administration

According to our surveys,[46,48] approximately 40%–60% of users omit at least some portion of the CAPE-V protocol. In fact, Lodhavia and Kempster[46] reported that "no single component of the CAPE-V was administered by all of the survey respondents 81% to 100% of the time" (p. 4). A few users eliminated the collection of one or both of the sustained vowels, used only a subset of the six sentences provided, provided a different prompt, or skipped the extemporaneous speech task entirely, opting instead to rate connected speech from conversation during the course of the voice evaluation. Nearly half of the respondents (47%) indicated they never use the specifically worded instructions provided in the protocol. On the other hand, adherence to the order of the tasks prescribed in the protocol is quite good, with nearly all (94%) respondents claiming to follow it at least sometimes.[46]

The CAPE-V committee intended the stimuli to be recorded and rated upon playback, providing specifics for doing so.[1] However, this procedure is not typically followed. In fact, adherence with this instruction may have decreased over time, given that 76.5% of the respondents surveyed in 2015[46] but only 16% of those surveyed in 2023[48] indicated that they always record the voice. Likewise, 6% of those surveyed in 2015 but 45% of the 2023 respondents reported that they never record the voice.

CAPE-V ratings and scoring

Although the vast majority of CAPE-V users rate the specified parameters of voice quality, few follow the rating and scoring procedures as prescribed in the published instructions.[1] All of the SLPs from the two surveys and one observational study rated overall severity of voice quality, and 82%-85% also rated roughness, breathiness, and strain.[46–48] A smaller majority (55%–75%) of clinicians completed ratings for pitch and loudness. The reasons for skipping these items, provided in open-ended comments, include perceived relevance, perceived validity, and efficiency[46–48]; some users commented that they measure acoustic correlates of pitch and loudness instead.

A scoring modification that SLPs frequently reported using is the estimation of ratings on each VAS rather than measuring the actual distance of tick marks along the lines. This strategy was reported by 77% of respondents in the survey by Lodhavia and Kempster.[46] Reasons provided generally were attributed to convenience and saving time but also included the belief that the auditory perception of voice quality was not that precise and that small differences were not meaningful.

The fidelity of following instructions for adding tick marks according to each of the three primary tasks was notably low. Only 17.6% of survey respondents reported that they always used this option, and 29% never did.[46] Nagle[47] observed that only one out of 20 SLPs marked two tasks differentially, and it was only for one voice-quality parameter for one speaker.

Certain judgments within the CAPE-V have proven to be difficult to standardize, such as the rating of "consistent" or "intermittent." The protocol instructions specified that consistency or intermittency should be marked if the attribute was or was not always present, respectively, within

and between tasks. The surveys and observations indicated that 65%–85% of users rated consistency at least sometimes but not necessarily in a systematic way. Lodhavia and Kempster[46] found that only 37% of respondents claimed to consistently mark "consistent" or "intermittent" for every perceptual dimension, and 15% of those observed by Nagle[47] never did. The majority (54%) of respondents surveyed by Lodhavia and Kempster indicated that they only noted consistency if a sample sounded inconsistent or intermittent. As such, if consistency is unmarked on the form, it is unclear whether voice quality attributes were consistently present or the examiner simply neglected to complete this item.

Contributing to the problem is confusion over how to interpret intermittency. For example, if a patient was judged to have a degree of roughness in extemporaneous speech productions but only at the ends of declarative sentences, and this pattern was consistent throughout the sampled speech, should this be marked as "consistent" or "intermittent?" Or, should the voice of a patient with spasmodic dysphonia be designated as "consistent" because voice breaks are apparent throughout the examination or "intermittent" because they do not occur at regular intervals? In addition, does intermittency affect different parameters differently? Although relatively few users report routinely scoring consistency, this characteristic of an individual's voice production is important to document for diagnostic purposes.

Audio recordings
The original authors of the CAPE-V specifically required that ratings be based on audio recordings obtained under specified conditions, not on live voice productions. Regardless, nearly half of users surveyed recently reported that they always administer and score the CAPE-V live, with the client/patient in front of the them.[48] Other clinicians reported completing voice evaluations in a telehealth format. Even if they record the voice samples, the samples are most likely recorded under a wide range of conditions (ie, various microphones and recording devices, mouth-to-microphone distances, adjusted volume controls). Given the current variability in the administration of this assessment, we contend that the CAPE-V should be scored based on listening to audio recordings. This allows clinicians to listen to a voice sample more than once before making a rating and to directly compare performance across tasks and across time. Ideally, repeated assessments of the same patient would use the same recording conditions and follow the same procedures.

Measuring and reporting results
The CAPE-V was developed to be measured as a ratio of the VAS. Users were directed to confirm that the VAS lines were 100 mm in length so that the number associated with a rating would be the distance (in mm) from the left end of the line. If the lines were not exactly 100 mm, the result

could be corrected proportionately (ie, dividing the distance of the tick mark from the left of the line by the entire length of the line). Electronic administration of the CAPE-V would obviously eliminate this measurement problem; an electronic scale might not measure 100 mm, but a score could still be reported in terms of the proportion of the scale to the left of the mark.

Some CAPE-V users have commented that they estimate the ratings, usually in multiples of 5 or 10.[47,48] In the survey conducted in Fall 2023, one participant noted that their perception is not refined enough to discriminate voice quality in units of 1/100, and another reported "estimating in 10s or categories."[48] This participant added: "Just because a VAS theoretically makes subjective ratings "parametric" isn't a good enough reason for me to use it."

CAPE-V Instructions: Efficiency and Technology
The original CAPE-V was intended to be used in reproducible, paper format. Although it is still a viable assessment/documentation format, paper files are considered outdated and impractical, especially with the transformation across healthcare centers and clinics to electronic medical records.[51,52] Consequently, it seems reasonable to encourage the development of electronic formats of the CAPE-V for clinical use, including at bedside. One such format currently exists in software developed by PentaxMedical called the iCAPE-V. This program can be bundled with other programs developed and sold by PentaxMedical for use with their Computerized Speech Lab system. Another version that can be used for training and clinical use is by Calaf.[53] In addition, according to our survey results,[48] locally developed programs are in limited use at certain facilities.

There are many advantages to an easily accessible electronic format. Electronic administration could provide standardized instructions for the clinician and clients, assist in digitizing and saving audio recordings, automatically score each section of the assessment, and generate a templated report that could be uploaded into electronic records.

Another advantage to full digital administration and scoring of an updated CAPE-V is that clinicians would be able to more easily examine relationships among the various components of a full voice assessment. This would promote full clinical review and comparison of the various components of a complete evaluation and would better substantiate the validity of a thorough and standard multidimensional evaluation. It would also improve inter-clinician and inter-clinic collaboration for the purposes of research and tracking patient outcomes.

REVISION OF THE CAPE-V
The issues raised in this review lead to the obvious conclusion that revisions to the original CAPE-V are warranted. In this section, we list and describe the modifications we implemented to create the CAPE-V – Revised (CAPE-Vr), displayed in Figure 3. Modifications are based on the premises that (a) the aims of the original CAPE-V committee

Consensus Auditory-Perceptual Evaluation of Voice – Revised (CAPE-Vr)

Name/ID: _____   **Gender:** _____   **Age:** _____

Examiner: _____   **Date:** _____

Recording Conditions

Audio recorded: Yes / No   In person / Virtual   Environment: Clinic room / Sound booth / Bedside / Home

Recording device/platform: _____   Mouth-to-mic (cm): _____

Stimuli

Vowels: /ɑ/ and /i/. Sustain each for 3 – 5 seconds; one or more productions in typical speaking voice.

Sentences:
a. The blue spot is on the key again.   d. I eat eggs every evening.
b. He helped her hurry home.   e. My mama makes lemon muffins.   ☐ *Check if the examiner*
c. We were away a year ago.   f. Papa took a piece of the cake.   *modeled the sentences.*

Extemporaneous Speech: "Tell me about a place you have gone or would like to go."

Reading Passage (optional): Specify: _____

Rating Conditions: Live voice / Recorded voice   Headphones / Speakers   Auditory anchors: Yes / No

Rater: _____   Date: _____   Number of times sample was played: ____

	Normal	Extreme
Overall Severity		____ /100
Roughness		____ /100
Breathiness		____ /100
Strain		____ /100
		____ /100

Pitch: Normal   Low   High   Comment: _____

Loudness: Normal   Quiet   Loud   Comment: _____

Resonance: Normal   Front   Back   Comment: _____

Nasality: Normal   Hyponasal   Hypernasal   Comment: _____

Inconsistencies: None / Present (describe):

Vowels: _____

Sentences: _____

Extemporaneous speech: _____

Instabilities: aphonic break   pitch break   pitch instability   spasm   tremor   other: _____

Additional features: aphonia   asthenia   diplophonia   falsetto   fry   hard glottal attack   wet/gurgly

other: _____

Overall Impression: _____

_____

FIGURE 3. The Consensus Auditory-Perceptual Evaluation of Voice – Revised (CAPE-Vr) form. For a reproducible copy that maintains proper dimensions, please see the supplementary material that accompanies the online version of this article.

are retained and respected; (b) modifications are supported by available research findings as much as possible; (c) actual usage of the CAPE-V by voice-focused clinicians is considered and incorporated as appropriate; and (d) the instructions and protocol are simplified and clarified. In addition to minor alterations, such as moving the space for examiner name to the top of the form and adding spaces for the gender and age of the patient, we made the following changes to the protocol and form.

Modification 1: Eliminated the textual severity markers below the VASs on the rating form

Despite the common use of the modifiers "mild," "moderate," and "severe" in the clinical assessment of voice, their appearance on the CAPE-V form has little to offer in terms of measurement sensitivity or clinical accuracy, especially given their differing placements on the two CAPE-V forms in circulation. For this reason, and in the interest of using a true (unlabeled) VAS,[54,55] we have removed the severity labels beneath the lines for the VASs in the CAPE-Vr. To remind raters of the direction of the VASs, we placed the words "normal" and "extreme" above the set of lines on the left and right sides, respectively.

Modification 2: Modified selected tasks and stimuli
Vowels

The original CAPE-V instructed users to obtain three repetitions each of two vowels. In the interest of time, the CAPE-Vr protocol suggests a single production of each vowel /ɑ/[b] and /i/, provided that the production is typical of the examinee's speaking voice in terms of pitch, loudness, and quality. Examiners determine whether productions are typical and can provide cues and implement strategies to ensure valid productions. Modeling is discouraged to avoid imitating the examiner's vocal pitch and quality.

Sentences

The CAPE-Vr modifies three of the sentence stimuli. Two of the original sentences provided with the CAPE-V should be revised as recommended by Zraick et al[3] and endorsed by multiple survey respondents. The sentence "How hard did he hit him?" contains aggressive language, and "We eat eggs every Easter" contains a religious reference, which users have reported to be offensive or not representative of the general population. Zraick et al suggested replacing them with "He helped Hannah hurry home" and "We eat eggs every evening." One of us has been using these sentences clinically for about 10 years and has found them to be easy to implement and just as interpretable as the original sentences. An additional modification to the latter sentence of substituting "We" with "I" gives another opportunity to observe whether examinees produce hard glottal attacks in their speech. To be culturally neutral, we

avoided using proper names (ie, *her* was preferred to *Hannah*). In sum, the first two sentence modifications of the CAPE-Vr are the introduction of "I eat eggs every evening" (sentence *d*) and "He helped her hurry home" (sentence *b*) to observe the effects of adductory and abductory laryngeal articulatory gestures, respectively.

The third sentence modification pertains to the sentence weighted with voiceless stop consonants followed by /i/, "Peter will keep at the peak," which some users identified as being awkward and often misread. This sentence was intended to reveal hypernasality by loading it with high oral-pressure consonants. In fact, fricatives[56,57] are equally, if not more, sensitive than stops in detecting hypernasality. Vowel context is less discriminatory than consonant class,[56] although /i/ is generally considered to provide the best context for revealing hypernasality[58,59] and is associated with greater velopharyngeal closing force compared to low vowels.[60] That said, it would be informative to observe if hypernasality were present in a variety of phonetic contexts that represent natural speech. Therefore, we composed a more natural, grammatically meaningful sentence with high-pressure consonants and several vowels that also avoids a proper name as a replacement: "Papa took a piece of the cake" (sentence *f*).

Although not modified, a clarification is needed for the nasal-consonant loaded sentence (sentence *e*). Kempster et al[1] stated in the body of their paper that the sentence is "My mamma makes lemon jam" although it is listed as "My mamma makes lemon muffins" on the form. We recommend using the sentence on the form ("…lemon muffins") for the sake of consistency and because recent research indicates this sentence is the most predictive CAPE-V speech task for ratings of overall severity.[61]

Finally, recognizing that clinicians frequently include paragraph-length material,[46,48] we included a space to indicate whether a reading passage is included. A common choice in voice research and in the clinic is the first paragraph of the Rainbow Passage. In fact, the second sentence extracted from this paragraph has proven to be a valid stimulus for differentiating dysphonic from nondysphonic voices using spectral/cepstral acoustic analysis,[62] which led to its recommended use in the Analysis of Dysphonia of Speech and Voice (PentaxMedical). Nonetheless, clinicians may choose to omit this optional task or select a different passage depending upon the patient's age, cognitive status, or other relevant variables.

Readers are reminded that the sentence stimuli were developed to represent a range of vocal functions and thus facilitate diagnostic testing across a variety of voice disorders. As non-English-speaking groups develop linguistically and culturally appropriate translations of the CAPE-Vr, they should keep these goals in mind.

Modification 3: Changed the extemporaneous speech prompt

The original prompts for eliciting spontaneous speech were "Tell me about your voice problem" or "Tell me how your

---

[b] The /a/ was changed to /ɑ/ to correctly reflect the typical low-back vowel used in American English.

voice is functioning."[1] These prompts have proven awkward for several reasons. Often, a version of these questions has already been posed and answered earlier in the assessment, so it becomes a redundant and time-wasting element. Some patients may be confused by the word "functioning." A key reason for avoiding these questions or variations thereof, however, is that they reveal information about the voice problem to a listener who may be tasked with providing an unbiased evaluation of a recorded version of the CAPE-V stimuli.[17,27] In addition, the prompt is not relevant to a person without vocal complaints who may be assessed with the CAPE-V as a baseline or screening assessment or participating in a research study.[3] Nonetheless, having a standard prompt to elicit and rate extemporaneous speech[c] is a worthwhile element to maintain in the CAPE-V assessment. Zraick et al[3] suggested "Tell me about your favorite holiday" or "Describe the neighborhood where you grew up." Another prompt reported in the literature is to ask the patient about their day or their daily routine,[11,12,21] but this can result in a list with repetitive ("sing-song") prosody. For the CAPE-Vr, we developed the prompt "Tell me about a place you have gone or would like to go." We selected this prompt after extensive discussion of a topic that would be unlikely to elicit a list, likely to generate rich linguistic content unrelated to voice concerns, and answerable by anyone regardless of religion, culture, or socioeconomic status. Beta testing of this prompt by one of us (NPS) in a clinical setting has resulted in acceptable speech samples that do not reveal the nature of the examinee's voice concerns.

Modification 4: Modified pitch and loudness scales to indicate direction but not extent of deviation from normal

Ratings for pitch and loudness on the CAPE-V required users to indicate the "Nature of the abnormality" with the intent of listing "too high" or "too low" on average. This response was often completed with consideration of the variability of the production rather than its average state or omitted altogether. Furthermore, the results were difficult to interpret if averaged across individuals or compared across time because there was no numerical way to differentiate too high from too low using the CAPE-V's scoring system. VAS ratings of pitch and loudness are rare in the literature, and users have reported using easily available acoustic correlates instead. In the CAPE-Vr, we have replaced VASs with categorical choices of normal, low, and high for pitch and normal, quiet, and loud for loudness. Space is also provided for additional comments for these parameters.

Modification 5: Added terms for rating resonance and nasality

The section on the original CAPE-V for commenting about resonance alluded to velopharyngeal functioning, in line with the high-pressure-consonant sentence included in the list of stimuli (sentence *f*). Clinically, however, resonance for patients with voice disorders usually focuses on the placement of perceived oral resonance; in fact, moving the focus of resonance from back to front is the basis for some voice therapy approaches.[63,64] Descriptions of hyponasality and hypernasality reflect the patency of nasal air passages and the functioning of the velopharyngeal mechanism, respectively. To more completely capture these features of oral and nasal resonance, the CAPE-Vr provides the opportunity to rate resonance as normal, front, or back, and nasality as normal, hyponasal, or hypernasal, with space to add comments for each.

Modification 6: Simplified and clarified observations of consistency

To accommodate the diagnostic value of observations of consistency in voice quality and to improve the efficiency of test administration, the CAPE-Vr provides space to report perceived inconsistencies according to the task. Rather than marking each vocal attribute for consistency and separately rating each attribute for each task, the CAPE-Vr provides a separate section in which raters can specify that there are no notable inconsistencies (by circling "None") or describe inconsistencies according to task. In addition, specific voice characteristics that reflect various types of vocal instability are listed alphabetically for ease of selection as appropriate. These are intended to add specificity to the term "pitch instability" that was included as an example in the list of "additional features" on the CAPE-V. The remaining original terms are now listed alphabetically on the CAPE-Vr form for selection, with the addition of "hard glottal attack" to provide examiners with an option for their observations on the sentence loaded with vowel-initial words (sentence *d*). As in the original CAPE-V, room is available to list features that are not provided in the fixed list.

Modification 7: Added space to write overall impression

Two lines are provided at the bottom of the CAPE-Vr form for clinicians to write their overall impression of the voice. This provides an opportunity to describe the overall severity of the voice disorder, list prominent voice features, and otherwise note memorable aspects of the perceptual evaluation.

Modification 8: Added sections about recording and analysis conditions

A final change to the CAPE-Vr form is the addition of sections on recording conditions and rating conditions. Recognizing that many voice assessments are now conducted virtually and may not be recorded, and that the

---

[c] Users have pointed out that the speech in this context is not spontaneous because it is prompted with a topic cue. Therefore, we have changed the terminology from *spontaneous* to *extemporaneous*.

environment and equipment may affect results, the Recording Conditions section prompts the examiner to note whether or not the voice was recorded, and if so, the nature of the environment and equipment used for recording purposes. A separate section for rating conditions includes space to indicate the identity of the rater (in case it differs from the examiner), when the ratings were completed, and other aspects of the rating methodology. These include the nature of the voice samples and playback conditions, the use of established examples of voices to represent different qualities and severities as auditory anchors, and the number of times the rater played the samples. The rationale for this modification is to provide consistent documentation of voice recording and rating conditions to be able to replicate conditions when evaluating voices over time and to improve the communication and transferability of findings across clinicians and researchers.

Modification 9: Updated instructions for administration

We have updated the instructions to accompany the CAPE-Vr form (Table 1; also provided as supplementary material). The information included in the new instructions is intended to maintain the original purpose of the instrument, as well as to clarify certain aspects of the protocol and rating form. These changes make this instrument more user friendly and reproducible across individuals and environments.

EXAMPLE OF CAPE-Vr IMPLEMENTATION

To illustrate the administration and scoring of the CAPE-Vr, we collected a recording of the new CAPE-Vr stimuli from a colleague familiar with the CAPE-V and skilled in feigning disordered voices. Recordings were made in a quiet room using an MXL USB unidirectional condenser

---

TABLE 1.
CAPE-Vr Protocol: Instructions for Administration and Scoring

Demographics
Complete identifying information about the examinee (Name or ID, Gender, Age), examiner (Name), and date of recording as indicated at the top of the form.

Recording Conditions
Seat the examinee comfortably in a quiet environment. Place a microphone at a fixed mouth-to-microphone distance and *audio record the voice and speech stimuli*. Note on the form whether the stimuli were audio recorded or not; whether the examination was in person or virtual; the room environment; the recording device and/or platform; and the mouth-to-microphone distance.

Tasks and Stimuli
The three primary tasks for the CAPE-Vr may be completed in any order.
Vowels. Instruct the examinee to say the *vowels /a/ and /i/* using a typical speaking voice. Each vowel should be *sustained for 3-5 seconds*. A single trial of each vowel is adequate if the production sounds representative of that individual's speaking voice. Modeling is discouraged to avoid imitation of the clinician's pitch and voice quality.
Sentences. Instruct the examinee to *read the following sentences aloud* in their typical speaking voice. The sentences can be printed out in large, easy-to-read font. If the individual has difficulty reading, the clinician may model the sentences; in this case, check the box provided to the right of the sentences on the CAPE-Vr form.
The sentences (and their primary features of interest) follow:
    a. The blue spot is on the key again. (English corner vowels)
    b. He helped her hurry home. (Word-initial /h/)
    c. We were away a year ago. (All voiced phonemes)
    d. I eat eggs every evening. (Vowel-initial words)
    e. My mama makes lemon muffins. (Nasal consonants)
    f. Papa took a piece of the cake. (High-pressure consonants)

Extemporaneous speech. Elicit *at least 20 seconds of natural speech* with the prompt: "Tell me about a place you have gone or would like to go." Another prompt that elicits content unrelated to voice use is also acceptable.
Reading passage (optional). Indicate the reading passage if one is included as part of the auditory-perceptual evaluation of voice.

Rating Conditions
Ratings are expected to be based on recordings of the voice, but the rater can indicate if they were completed on live voice in real time instead. When rating voice recordings, listen to the stimuli as many times as desired and document the number of repetitions. Also indicate: the use headphones or speakers; the use of standard samples of disordered voices as auditory anchors (ie, reference samples); the identity of the rater; and the date.

Voice Characteristics and Ratings

Attributes Rated along Visual Analog Scales (VASs)

TABLE 1 (*Continued*)

The salient perceptual vocal attributes included in the CAPE-V and CAPE-Vr were identified by the original consensus
  committee authors as commonly used and easily understood: Overall Severity, Roughness, Breathiness, and Strain.
  These vocal attributes are generally defined as follows:
Overall Severity: global, integrated impression of deviation from normal voice
Roughness: perceived irregularity in the voicing source
Breathiness: perceived air escape in the voice
Strain: perceived vocal effort, tension, or press
Each of these perceptual attributes is accompanied by a 100-mm line forming a VAS. One blank VAS is included on the
  form if the examiner prefers to rate an additional attribute on a continuous scale. The words "Normal" and "Extreme"
  appear above the lines on the left and right sides, respectively, to indicate the direction of the perceptual ratings.
The examiner marks the degree of perceived deviance for each attribute with a small vertical line (aka a 'tick mark'). The
  examiner may mark each VAS at any location.

Scoring VASs. After marking each of the VASs, measure the distance (in mm) from the left end of the line to the tick mark.
  Write the value in the blank space to the right of the line. Confirm that the lines are 100 mm long; if not, cross out 100 and
  insert the actual length of the line. Corrections can be made by dividing the distance measured by the length of the line
  and multiplying that result by 100.

Attributes Rated Descriptively
Options are provided for auditory-perceptual judgments of pitch, loudness, resonance, and nasality, inconsistencies,
  instabilities, and additional features, with room for examiner comments. Examiners may select one or more options per
  attribute.
Pitch: perceived average pitch of the voice, as Normal, Low, or High. Pitch variability may be noted separately under
  Inconsistencies or Instabilities.
Loudness: perceived average loudness or sound level of the voice, as Normal, Quiet, or Loud. Loudness variability may be
  noted separately under Inconsistencies or Instabilities.
Resonance: perceived focus of the sound within the oral cavity, as Normal, Front (forward or "in the mask"), and Back
  (pharyngeal or "throaty").
Nasality: perceived balance of oral and nasal resonance reflecting patency of the nasal passageways and velopharyngeal
  function, as Normal, Hyponasal, or Hypernasal.
Inconsistencies: Indicate and describe inconsistencies of voice according to task. If there are no notable inconsistencies,
  circle None. If there are, circle Present and provide a description of the inconsistencies.
Instabilities: If relevant, circle one or more categorical descriptors of vocal instabilities from the options provided or add
  another term to indicate vocal instabilities.
Additional features: If relevant, select one or more descriptors from the options provided or list alternate descriptors.
Overall impression: State the overall severity of the voice problem and describe the voice in a few words or phrases.

microphone placed 10 cm from the lips. Based on the di-
gitally recorded file, we each independently completed the
CAPE-Vr form. We then reviewed how we approached the
rating task and our resultant independent ratings.

Figure 4 shows the form as completed by the rater who
provided the median score for overall severity. We did not
attempt to reach a consensus and coalesce our ratings for
the purposes of reliability of measurement. Rather, our
goal was to ensure that we could follow the processes
outlined in the CAPE-Vr and that our final adjustments
would maximize fidelity and represent a doable, easily
understood, standard approach.

The VAS lines were confirmed to be 100 mm long on the
printed form. As shown, overall severity was rated 77/100,
roughness 83/100, breathiness 35/100, and strain 46/100.
The extra VAS option was unused. Pitch was judged to be
lower than normal and oral resonance was perceived as
back. Loudness and nasality were considered to be normal.
Inconsistency was noted only for vowels with pitch in-
stability at the onset of /ɑ/. The rater also noted fry and

possible diplophonia. Finally, she added a comment at the
bottom of the form with her overall impression and to
highlight prominent attributes of the voice.

Each of our experiences using the CAPE-Vr revealed
that the new form was clear, thorough without being
overwhelming, and logically organized. The completed
example displayed in Figure 4 uses the final version of the
form. We include this form for teaching and presenting
purposes to show how the CAPE-Vr would optimally be
used for the auditory-perceptual evaluation of a recorded
voice.

SUMMARY

The CAPE-V has had a significant impact on both the
clinical assessment of voice quality and in many research
endeavors internationally for more than 20 years. In this
article, we highlighted issues that have arisen over time
with the administration, stimuli, rating form, scoring, and
instructions accompanying the 2002 online beta-version

Consensus Auditory-Perceptual Evaluation of Voice – Revised (CAPE-Vr)

Name/ID: _D. B._   Gender: _M_   Age: _53_

Examiner: _KFN_   Date: _8/13/24_

Recording Conditions

Audio recorded (Yes) No   (In person) / Virtual   Environment: (Clinic room) / Sound booth / Bedside / Home

Recording device/platform: _MXL Condensor mic – unidirectional_   Mouth-to-mic (cm): _10_

Stimuli

Vowels: /ɑ/ and /i/. Sustain each for 3 – 5 seconds; one or more productions in typical speaking voice.

Sentences:

a. The blue spot is on the key again.   d. I eat eggs every evening.

b. He helped her hurry home.   e. My mama makes lemon muffins.   ☐ *Check if the examiner*

c. We were away a year ago.   f. Papa took a piece of the cake.   *modeled the sentences.*

Extemporaneous Speech: "Tell me about a place you have gone or would like to go."

Reading Passage (optional): Specify: _Rainbow Passage_

Rating Conditions: Live voice / (Recorded voice)   Headphones / (Speakers)   Auditory anchors: Yes / (No)

Rater: _GBK_   Date: _12/30/24_   Number of times sample was played: _4_

	Normal	Extreme	
Overall Severity			_77_ /100
Roughness			_83_ /100
Breathiness			_35_ /100
Strain			_44_ /100
			___ /100

Pitch:	Normal	(Low)	High	Comment: _____
Loudness:	(Normal)	Quiet	Loud	Comment: _____
Resonance:	Normal	Front	(Back)	Comment: _____
Nasality:	(Normal)	Hyponasal	Hypernasal	Comment: _possibly mildly "hypo"_

Inconsistencies: None / (Present) (describe):

Vowels: _/a/ unstable at onset_

Sentences: _____

Extemporaneous speech: _____

Instabilities: aphonic break   pitch break   (pitch instability)   spasm   tremor   other: _____

Additional features: aphonia   asthenia   (diplophonia?) / falsetto   (fry)   hard glottal attack   wet/gurgly

other: _____

Overall Impression: _Moderate-severe dysphonia c̄ low pitch,_
"throaty" resonance, & roughness.

FIGURE 4. Example of a completed CAPE-Vr form.

tool and the 2009 publication of the original CAPE-V. Based on published literature, available survey data, clinical experience, and practical considerations, this paper provides the rationale and development for a revised CAPE-V, the *CAPE-Vr*.

The importance of maintaining fidelity to the CAPE-Vr protocol—or to any set of assessment instructions—is required to maintain the validity of the CAPE-Vr. As we train the next generation of speech-language pathologists, it is important that the procedures we espouse are clear, consistent, and well-founded in theory, research, and experience. The application of implementation-science strategies and methods can direct future investigations to ascertain whether clinicians more easily and naturally maintain fidelity to the CAPE-Vr than they did to the CAPE-V. It is our expectation and our hope that fidelity to the CAPE-Vr will facilitate evidence-based practices for clinicians assessing individuals with voice disorders.

Data Availability Statement

Data sharing is not applicable to this article as no data sets were generated or analyzed for the current work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Author Contributions

All authors contributed equally to the conceptualization, development, drafting and finalization of this work.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the official position of the Department of Defense or the U.S. Government.

Appendix A. Supporting Information

A supplementary file including the CAPE-Vr form and protocol with this article can be found in the online version at doi:10.1016/j.jvoice.2025.01.022.

References

1. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol.* 2009;18:124–132.
2. Karnell M, Melton S, Childes J, Coleman T, Dailey S, Hoffman H. Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *J Voice.* 2007;21:576–590.
3. 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;20:14–22.
4. Hirano M. "GRBAS" scale for evaluating the hoarse voice and frequency range of phonation. In: Hirano M, ed. Wien: Springer-Verlag; 1981:88–89. Clinical Examination of Voice; 5.
5. Dejonckere PH, Obbens C, Moor GM de, Wieneke GH. Perceptual evaluation of dysphonia: reliability and relevance. *Folia Phoniatr Logop.* 1993;45:76–83. https://doi.org/10.1159/000266220.
6. Webb A, Carding P, Deary I, MacKenzie K, Steen N, Wilson J. The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Otorhinolaryngol.* 2004;261:429–434.
7. Wuyts F, De Bodt M, Van de Heyning P. Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J Voice.* 1999;13:508–517.
8. Mahalingam S, Venkatraman Y, Boominathan P. Cross-cultural adaptation and validation of consensus auditory perceptual evaluation of voice (CAPE-V): a systematic review. *J Voice.* 2024;38:630–640. https://doi.org/10.1016/j.jvoice.2021.10.022.
9. Ryan MA, Brodsky MB, Blumin JH, et al. Twenty-one for 2021: the most influential papers in laryngology since 2000. *Laryngoscope.* 2022;132:406–412. https://doi.org/10.1002/lary.29781.
10. Salgado S, Schils SA, Childes JM, Crino C, Palmer AD. Current practices in the assessment of voice: a comparison of providers across different clinical settings. *J Voice.* 2024. https://doi.org/10.1016/j.jvoice.2024.08.007.
11. Joshi A, Baheti I, Angadi V. Cultural and linguistic adaptation of the consensus auditory-perceptual evaluation of voice (CAPE-V) into hindi. *J Speech Lang Hear Res.* 2020;63:3974–3981. https://doi.org/10.1044/2020_JSLHR-20-00348.
12. Chen Z, Fang R, Zhang Y, et al. The Mandarin Version of the consensus auditory-perceptual evaluation of voice (CAPE-V) and its reliability. *J Speech Lang Hear Res.* 2018;13:1–7. https://doi.org/10.1044/2018_JSLHR-S-17-0386.Published online September.
13. Gunjawate DR, Ravi R, Bhagavan S. Reliability and validity of the Kannada version of the consensus auditory-perceptual evaluation of voice. *J Speech Lang Hear Res.* 2020;63:385–392. https://doi.org/10.1044/2019_JSLHR-19-00020.
14. Mozzanica F, Ginocchio D, Borghi E, Bachmann C, Schindler A. Reliability and validity of the Italian version of the consensus auditory-perceptual evaluation of voice (CAPE-V). *Folia Phoniatr Logop.* 2013;65:257–265. https://doi.org/10.1159/000356479.

15. Pommée T, Mbagira D, Morsomme D. French-language adaptation of the consensus auditory-perceptual evaluation of voice (CAPE-V). *J Voice.* 2024. https://doi.org/10.1016/j.jvoice.2024.03.011.

16. Núñez-Batalla F, Morato-Galán M, García-López I, Avila-Menéndez A. Validation of the Spanish adaptation of the consensus auditory-perceptual evaluation of voice (CAPE-V). *Acta Otorrinolaringol Esp.* 2014;66:249–257. https://doi.org/10.1016/j.otorri.2014.07.007.

17. Calaf N, Garcia -Quintana D. Development and validation of the bilingual Catalan/Spanish cross-cultural adaptation of the consensus auditory-perceptual evaluation of voice. *J Speech Lang Hear Res.* 2024;67:1072–1089. https://doi.org/10.1044/2024_JSLHR-23-00536.

18. Kondo K, Mizuta M, Kawai Y, et al. Development and validation of the Japanese version of the consensus auditory-perceptual evaluation of voice. *J Speech Lang Hear Res.* 2021;64:4754–4761. https://doi.org/10.1044/2021_JSLHR-21-00269.

19. Venkatraman Y, Mahalingam S, Boominathan P. Development and validation of sentences in tamil for psychoacoustic evaluation of voice using the consensus auditory-perceptual evaluation of voice. *J Speech Lang Hear Res.* 2022;65:4539–4556. https://doi.org/10.1044/2022_JSLHR-22-00169.

20. Mossadeq NM, Khairuddin KAM, Zakaria MN. Cross-cultural adaptation of the consensus auditory-perceptual evaluation of voice (CAPE-V) into Malay: a validity study. *J Voice.* 2022;38:1527.e27–1527.e35. https://doi.org/10.1016/j.jvoice.2022.05.018.

21. Ertan-Schlüter E, Demirhan E, Ünsal EM, Tadıhan-Özkan E. The Turkish version of the consensus auditory-perceptual evaluation of voice (CAPE-V): a reliability and validity study. *J Voice.* 2020;34:965.e13–965.e22. https://doi.org/10.1016/j.jvoice.2019.05.014.

22. Özcebe E, Aydınlı FE, Tiğrak TK, İncebay Ö, Yilmaz T. Reliability and validity of the Turkish version of the consensus auditory-perceptual evaluation of voice (CAPE-V). *J Voice.* 2019;33:382.e1–382.e10. https://doi.org/10.1016/j.jvoice.2017.11.013.

23. Almeida SC, de, Mendes AP, Kempster GB. The consensus auditory-perceptual evaluation of voice (CAPE-V) psychometric characteristics: II European Portuguese Version (II EP CAPE-V). *J Voice.* 2019;33:582.e5–582.e13. https://doi.org/10.1016/j.jvoice.2018.02.013.

24. Jesus LM, Barney A, Couto P, Vilarinho H, Correia A. Voice quality evaluation using CAPE-V and GRBAS in European Portuguese. In Proceedings of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA). 2009; Florence, Italy: 61-64.

25. Behlau M, Rocha B, Englert M, Madazio G. Validation of the Brazilian Portuguese CAPE-V Instrument—Br CAPE-V for auditory-perceptual analysis. *J Voice.* 2022;36:586.e15–586.e20. https://doi.org/10.1016/j.jvoice.2020.07.007.

26. Helou L, Solomon N, Henry L, Coppit G, Howard R, Stojadinovic A. The role of listener experience on consensus auditory-perceptual evaluation of voice (CAPE-V) ratings of postthyroidectomy voice. *Am J Speech Lang Pathol.* 2010;19:248–258.

27. Solomon NP, Helou LB, Stojadinovic A. Clinical versus laboratory ratings of voice using the CAPE-V. *J Voice.* 2011;25:e7–e14. https://doi.org/10.1016/j.jvoice.2009.10.007.

28. ASHA. Consensus auditory-perceptual evaluation of voice (CAPE-V) ASHA Special Interest Division 3, Voice and Voice Disorders. Published online 2002. Available at: http://www.asha.org/uploadedFiles/ASHA/SIG/03/CAPE-V-Procedures.pdf. Accessed August 4, 2009.

29. Kempster G. CAPE-V: development and future direction. *Perspect Voice Voice Disord.* 2007;17:11–13. https://doi.org/10.1044/vvd17.2.11.

30. Connor NP, Bless D, Dardis C, Vinney L, DoIT Engage Project Team. Simulations: consensus auditory-perceptual evaluation of voice [CAPE-V]. Voice disorders: simulations and games. 2008. Available at: https://csd.wisc.edu/slpgames/sims.html. Accessed August 20, 2024.

31. Boone D, McFarlane S, von Berg S, Zraick RI. *The Voice and Voice Therapy.* 10th ed.., London, U.K: Pearson,; 2020.

32. Ferrand CT. *Voice Disorders: Scope of Theory and Practice.* 2nd ed.., London, U.K.: Pearson,; 2019.

33. Stemple JC, Roy N, Klaben BK. *Clinical Voice Pathology: Theory and Management.* 7th ed.., San Diego, CA: Plural Publishing,; 2026.

34. Awan S, Lawson L. The effect of anchor modality on the reliability of vocal severity ratings. *J Voice.* 2009;23:341–352.

35. Kelchner LN, Brehm SB, Weinrich B, et al. Perceptual evaluation of severe pediatric voice disorders: rater reliability using the consensus auditory perceptual evaluation of voice. *J Voice.* 2010;24:441–449. https://doi.org/10.1016/j.jvoice.2008.09.004.

36. Stojadinovic A, Henry LR, Howard RS, et al. Prospective trial of voice outcomes after thyroidectomy: evaluation of patient-reported and clinician-determined voice assessments in identifying post-thyroidectomy dysphonia. *Surgery.* 2008;143:732–742.

37. Braden MN, Johns MMI, Klein AM, Delgaudio JM, Gilman M, Hapner ER. Assessing the effectiveness of botulinum toxin injections for adductor spasmodic dysphonia: clinician and patient perception. *J Voice.* 2010;24:242–249. https://doi.org/10.1016/j.jvoice.2008.08.003.

38. Sewall GK, Jiang J, Ford CN. Clinical evaluation of Parkinson's-related dysphonia. *Laryngoscope.* 2006;116:1740–1744. https://doi.org/10.1097/01.mlg.0000232537.58310.22.

39. Nagle KF. Challenges to CAPE-V as a standard. *Perspect ASHA Spec Interest Groups SIG 3.* 2016;1:47–53. https://doi.org/10.1044/persp1.SIG3.47.

40. Feinstein H, Daşdöğen Ü, Awan JA, Awan SN, Abbott KV. Comparative analysis of two methods of perceptual voice assessment. *J Voice.* 2023. https://doi.org/10.1016/j.jvoice.2023.01.005.

41. Walden PR. Perceptual voice qualities database (PVQD): database characteristics. *J Voice.* 2020;36:875.e15–875.e23. https://doi.org/10.1016/j.jvoice.2020.10.001.

42. Walden PR, Rau S. Individual voice dimensions' prediction of overall dysphonia severity on two auditory-perceptual scales. *J Speech Lang Hear Res.* 2022;65:2759–2777. https://doi.org/10.1044/2022_JSLHR-21-00689.

43. Kania E, Davis M, Walden P. Severity cutoffs for auditory-perceptual voice quality measurements. Poster presented at: The Voice Foundation Symposium: Care of the Professional Voice. June 2023; Philadelphia.

44. Nagle KF, Helou LB, Solomon NP, Eadie TL. Does the presence or location of graphic markers affect untrained listeners' ratings of severity of dysphonia. *J Voice.* 2014;28:469–475. https://doi.org/10.1016/j.jvoice.2013.12.011.

45. ASHA. Consensus auditory-perceptual evaluation of voice (CAPE-V) ASHA Special Interest Division 3, Voice and Voice Disorders. Published online 2009. Available at: https://www.asha.org/Form/CAPE-V-Success/. Accessed October 13, 2019.

46. Lodhavia A, Kempster GB. Fidelity to the consensus auditory-perceptual analysis of voice (CAPE-V): a pilot study. *J Voice.* 2024. https://doi.org/10.1016/j.jvoice.2023.12.009.

47. Nagle KF. Clinical use of the CAPE-V scales: agreement, reliability and notes on voice quality. *J Voice.* 2022. https://doi.org/10.1016/j.jvoice.2022.11.014.

48. Nagle KF, Kempster GB, Solomon NP. Survey of voice-focused speech-language pathologists' usage of the consensus auditory perceptual evaluation of voice (CAPE-V). *J Voice.* 2024. https://doi.org/10.1016/j.jvoice.2024.08.032.

49. Patel RR, Awan SN, Barkmeier-Kraemer J, et al. Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *Am J Speech Lang Pathol.* 2018;25:1–19. https://doi.org/10.1044/2018_AJSLP-17-0009. Published online June.

50. Roy N, Barkmeier-Kraemer J, Eadie T, et al. Evidence-based clinical voice assessment: a systematic review. *Am J Speech Lang Pathol.* 2013;22:212–226. https://doi.org/10.1044/1058-0360(2012/12-0014).

51. Boonstra A, Versluis A, Vos JFJ. Implementing electronic health records in hospitals: a systematic literature review. *BMC Health Serv Res.* 2014;14:370. https://doi.org/10.1186/1472-6963-14-370.

52. Nguyen L, Bellucci E, Nguyen LT. Electronic health records implementation: an evaluation of information system impact and contingency factors. *Int J Med Inf.* 2014;83:779–796. https://doi.org/10.1016/j.ijmedinf.2014.06.011.

53. Calaf N. All-Voiced [Web application]. 2024. Available at: https://www.all-voiced.com/. Accessed August 20, 2024.

54. Dexter F, Chestnut DH. Analysis of statistical tests to compare visual analog scale measurements among groups. *Anesthesiology.* 1995;82:896–902. https://doi.org/10.1097/00000542-199504000-00012.

55. Heller GZ, Manuguerra M, Chow R. How to analyze the visual analogue scale: myths, truths and clinical relevance. *Scand J Pain.* 2016;13:67–75. https://doi.org/10.1016/j.sjpain.2016.06.012.

56. Morris HL, Spriestersbach DC, Darley FL. An articulation test for assessing competency of velopharyngeal closure. *J Speech Hear Res.* 1961;4:48–55. https://doi.org/10.1044/jshr.0401.48.

57. Subtelny JD, Subtelny JD. Intelligibility and Associated Physiological factors of cleft palate speakers. *J Speech Hear Res.* 1959;2:353–360. https://doi.org/10.1044/jshr.0204.353.

58. Kataoka R, Zajac DJ, Mayo R, Lutz RW, Warren DW. The influence of acoustic and perceptual factors on perceived hypernasality in the vowel [i]: a preliminary study. *Folia Phoniatr Logop.* 2001;53:198–212. https://doi.org/10.1159/000052675.

59. Zraick R, Liss J. A comparison of equal-appearing interval scaling and direct magnitude estimation of nasal voice quality. *J Speech Lang Hear Res.* 2000;43:979–988.

60. Moon SJ, Lindblom B. Interaction between duration, context, and speaking style in English stressed vowels. *J Acoust Soc Am.* 1994;96:40–55.

61. Braun B, Incorvaia S, Walden P. Relationship of cepstral peak prominence and auditory-perceptual judgments of severity in varied phonetic contexts. Podium presented at: Voice Foundation Symposium: Care of the Professional Voice; June 2024; Philadelphia.

62. Watts CR, Awan SN. Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *J Speech Lang Hear Res.* 2011;54:1525–1537. https://doi.org/10.1044/1092-4388(2011/10-0209).

63. Stemple JC. A holistic approach to voice therapy. *Semin Speech Lang.* 2005;26:131–137. https://doi.org/10.1055/s-2005-871209.

64. Yiu EML, Lo MCM, Barrett EA. A systematic review of resonant voice therapy. *Int J Speech Lang Pathol.* 2017;19:17–29. https://doi.org/10.1080/17549507.2016.1226953.

65. Kempster GB, Nagle KF, Solomon NP. Development and rationale for the consensus auditory-perceptual evaluation of voice – revised (CAPE-Vr). *PsyArXiv.* 2024. https://doi.org/10.31234/osf.io/e84tn.