Survey of Voice-Focused Speech-Language Pathologists' Usage of the Consensus Auditory Perceptual Evaluation of Voice (CAPE-V)

[*]Kathleen F. Nagle, [†,1]Gail B. Kempster, and [‡]Nancy Pearl Solomon, *Nutley, New Jersey, †Chicago, Illinois, and ‡Bethesda, MD

Summary: Purpose. As part of the process of developing specific recommendations for modifying certain elements of the Consensus Auditory Perceptual Evaluation of Voice (CAPE-V) to promote end-user fidelity, the authors sought input from voice clinicians who regularly use the CAPE-V to assess voice quality.

Method. At an academic meeting focusing on voice disorders, we presented a poster briefly reviewing the CAPE-V protocol and describing several sources of variability that have been reported in its current use. Interested viewers were directed to a QR code linking to a brief, anonymous survey on how individuals currently use the CAPE-V and how they might improve it. A link to the survey was also distributed on the conference discussion board.

Results. Fifty-nine participants responded to the survey: 49 completed it. The median respondent reported 8 years of experience conducting voice evaluations, with 50% of their current practice in voice, and about eight voice evaluations per week. Key findings from this survey were that fewer than half of respondents reported audio recording any components of in-person or virtual voice evaluations, and that most respondents reported changing some aspect of the CAPE-V tasks and stimuli in practice.

Conclusion. This exploratory study revealed a wide range of idiosyncratic practices by clinicians when administering and scoring the CAPE-V. The findings support planned revisions to the CAPE-V protocol and form involving the tasks, stimuli, and rating procedures.

Key Words: CAPE-V–Auditory-perceptual–Evaluation–Fidelity.

The Consensus Auditory Perceptual Evaluation of Voice (CAPE-V) was introduced by ASHA in 2002 based on the work of a SIG-3 sponsored committee, and published by Kempster and colleagues in the *American Journal of Speech-Language Pathology* in 2009.[1] It was intended to provide a standard, brief, and reliable protocol for perceptually evaluating voice quality in both research and clinical settings. Since its initial release in 2002, the protocol has been translated and validated in numerous languages, and it is widely used among voice clinicians and researchers.[2,3] However, there is increasing evidence of confusion about some of the characteristics of the instrument and of the protocol itself. For example, clinicians indicated in a 2015 survey that although they generally elicited a vowel and extemporaneous speech as directed by the protocol, they did not always elicit the prescribed sentences; likewise, some reported that they never rate the parameters of Roughness, Breathiness, or Strain.[4] These findings are consistent with a more recent study of how 20 experienced voice clinicians reported using the CAPE-V.[5] As part of the process of developing specific recommendations for modifying certain elements of the CAPE-V protocol and for revising its documentation form, we sought input from a larger group of clinicians who use the CAPE-V regularly.

METHOD

This study was approved as exempt by the Seton Hall University Institutional Review Board (#2024-493).

Survey

We created an anonymous web-based survey (Qualtrics October 2023 version, Provo, UT) with five sections focused on how experienced clinicians elicit and rate voice stimuli when administering the CAPE-V (Appendix A). Section 1 included a few demographic questions; to keep the survey anonymous, we limited the number and type of questions related to individual participants' personal characteristics. Section 2 had questions about how participants conduct voice evaluations using the CAPE-V. In particular, we were interested in modality of the visit (ie, virtual or in-person); how CAPE-V ratings were captured (ie, paper vs electronic scales); and whether the voice productions were recorded. The instructions for the CAPE-V provide details about using a paper version of the form, including the advice to "verify that your paper copy has

Accepted for publication August 26, 2024.

Portions of this work were presented at the 2023 ASHA Convention in Boston, Massachusetts and the 53rd Annual Voice Foundation Symposium (2024) in Philadelphia, Pennsylvania.

From the *Seton Hall University, Nutley, New Jersey; †Rush University, Chicago, Illinois; and the ‡Walter Reed National Military Medical Center, Uniformed Services University of the Health Sciences, Bethesda, MD

[1] Retired..

Address correspondence and reprint requests to Kathleen F. Nagle, Department of Speech-Language Pathology, Seton Hall University, Interprofessional Health Sciences (IHS) Campus, I Building 123, 123 Metro Boulevard, Nutley, NJ 07110. E-mail: kathleen.nagle@shu.edu

Journal of Voice, Vol xx, No xx, pp. xxx–xxx

0892-1997

https://doi.org/10.1016/j.jvoice.2024.08.032

accurate 100-mm lines before reproducing the CAPE-V form"[1]; however, an electronic version of the CAPE-V form has been licensed by ASHA and is available as the iCAPE-V with the Computerized Speech Lab as of 2019 (Pentax Medical, Montvale, NJ).[a] The CAPE-V protocol also provides specific instructions to audiorecord the examinee's performance on the three production tasks (ie, vowels, sentences, and spontaneous speech).

Section 3 of the survey focused on how participants elicit the CAPE-V stimuli: of interest were which stimuli are elicited and what prompts are used to elicit them. Section 4 addressed how participants rate stimuli using the CAPE-V form and scales. No guidance is given in the CAPE-V instructions as to when clinicians should rate voice quality on the form apart from a direction to audiorecord the tasks and to complete all tasks before rating.

In Section 5, participants were asked to explain their reasoning if they skipped sections of the CAPE-V and what they would change about the CAPE-V if they could. Finally, participants were given the option to provide their email address to be contacted by the authors if they wished to share additional thoughts about the CAPE-V.

Before releasing the survey, we received feedback on its questions and wording from several voice clinicians who were not eligible to participate in the survey because they had been interviewed for a previous related study.[5] These pilot participants reported that it took 8 minutes or less to complete the survey.

Presentations

We presented the survey to attendees of the Fall Voice Conference in Washington, DC, in October 2023, first via QR code on the poster and flyers that were distributed to attendees, and then in a post to the online discussion board for the conference (ie, "Fall Voice Community: Eat, Breathe, Talk"). This post contained both the QR code and a link to the survey. The survey remained open for 25 days (10/19/23-11/13/23).

As shown in Appendix B, the recruitment poster very briefly described some reported sources of variability in the use of the CAPE-V protocol and form, specifically:

- Some clinicians record stimuli; others rate them only in real time
- Some CAPE-V stimuli and tasks are modified or skipped by some clinicians
- There are two versions of the CAPE-V form, with different labeling of the rating scales
- Not all clinicians mark consistency within or across stimuli
- Not all clinicians rate the main parameters of Breathiness, Roughness, and Strain

These points were followed by a QR code in the middle of the poster inviting viewers to take the 10-minute anonymous survey. The Discussion section of the poster contained draft recommendations for revisions to the CAPE-V protocol and form, along with future research questions such as whether the modality (and timing) of scoring CAPE-V parameters would make a difference in the ratings. The flyer distributed at the conference briefly stated the purpose of the survey, a statement about consent, and the QR code to access the survey. The discussion board post included the same information as the flyer.

Data analysis

All data from the survey are provided in the Supplementary Materials, including mean, standard deviation, median and range for continuous data, and frequency counts and proportions of categorical data. Text responses to these questions (ie, accompanying "Other" or "Please explain" choices) were assigned to categories and included in frequency counts as appropriate. For example, in response to Question 4 ("In what type of setting do you practice?"), the choice of "Other" with a response of "outpatient" was counted as a hospital setting. To enhance the readability of this paper, most of the numerical data are summarized in the text and presented as Supplementary Tables.

Narrative responses to open-ended questions were reviewed for thematic similarity. Each of the authors independently categorized responses to Questions 26 and 27 to identify themes and extracted quotes that were "illustrative, succinct, and representative" of the general patterns within the data (Lingard).[6] Responses to earlier questions that addressed Questions 26 or 27, regarding what sections of the CAPE-V they skip and what they would change about the CAPE-V if they could, were included in this qualitative analysis. Discrepant examples were also purposefully identified so that a diversity of reasoning was represented for skipping sections of the CAPE-V protocol or suggesting changes to it. A final set of main themes with exemplars was then developed by consensus discussion among the authors for each question.

RESULTS

Eighty-eight individuals clicked on the QR code (n = 42) or link (n = 46) to open the survey, and 59 participants proceeded beyond the informed-consent screen to respond to at least some of the questions. Of these, 55 responded to all of the questions in Sections 1-3, regarding the amount of time to elicit stimuli; 49 continued by responding to all of the questions in Section 4, regarding rating and scoring; and 46 completed the entire survey, including open-ended questions in Section 5.

Demographic information

Demographic information from Questions 1 to 5 is presented in Table 1. Most of the 49 participants who completed the survey reported working in a hospital setting,

[a] The CAPE-V instructions[1] state: "It is strongly recommended that for all rating sessions following the initial one, the clinician have a paper or electronic copy of the previous CAPE-V ratings available for comparison purposes."

TABLE 1.
Descriptive Statistics for Demographic Questions Related to Voice Practice

	Years of voice practice	Proportion of practice that is voice (%)	Number of voice evaluations/ week
Mean	10.3	60.3	9.5
SD	8.4	27.4	7.2
Median	8.0	50.0	8.0
Range	1-44	0-100	0-30

including inpatient, outpatient, acute, and long-term acute care. Fourteen reported working in a university clinic, 11 in a private clinic, and one in a skilled nursing facility. At least a few participants reported working in more than one setting, given that there were 71 responses to Question 4. About a third of participants (36%) reported having used the CAPE-V for 1-5 years, with 27% having used it for 5-10 years, and 29% for more than 10 years. Eight percent of respondents had been using the CAPE-V for less than a year.

As shown in Table 1, respondents had a wide range of experience in years and proportion of their practice being voice, with some reporting up to 30 voice evaluations per week. Reports of 0% and 0 voice evaluations per week were assumed to be from speech-language pathology (SLP) students.

All responses to Questions 6-28 are provided in Tables Q6-Q28 in Supplementary Materials.

Evaluation procedure

Of the 58 participants who responded to Question 6, 50 reported doing most or all voice evaluations in-person and six reported an even split; only two reported doing most or all evaluations virtually.

Based on Question 7 about administration modality, 68% of the 25 respondents who administer the CAPE-V in-person use the paper version of the CAPE-V scales. Five reported using a clinic-developed electronic version of the scales, and two use a web-based electronic version of the scales, presumably the iCAPE-V. One participant reported using "shorthand via Epic" to rate voice quality. Of the 25 clinicians who answered Question 8 about recording the voice samples, the majority (60%) audiorecord the stimuli; 16% also videorecord the patient. Forty percent reported not recording the evaluation at all.

Of the 46 participants who answered Question 9, regarding recording the voice when evaluating a patient virtually, 24% record audio only; 13% record both audio and video content; and 50% do not record the evaluation at all. The 13% who chose "Other" did not provide details.

Eliciting stimuli

It was not possible to establish how many participants responded to Questions 10-14, because they could make multiple choices; however, at least 46 responded to Question 10, regarding the tasks they typically elicit when administering the CAPE-V (Figure 1). The seven participants who chose "other" reported that they obtain productions of ascending and descending pitch glides, "therapy probe activities," reading passages, and /u/ for judging voice quality.

At least 41 participants responded to Question 11, regarding which of the CAPE-V sentences they use (Figure 2). Three of the participants who chose "other" use the standard CAPE-V sentences translated into another language (ie, Greek, Spanish, and Turkish; Table Q12 in Supplementary Materials). One participant indicated that

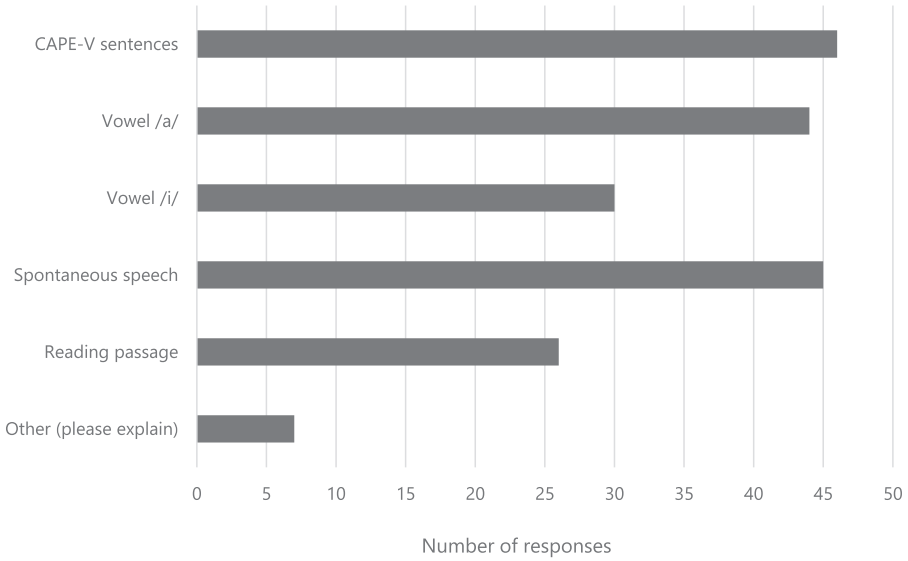

FIGURE 1. Frequency counts for use of specific components reported by participants in a typical administration of the CAPE-V protocol (Question 10).

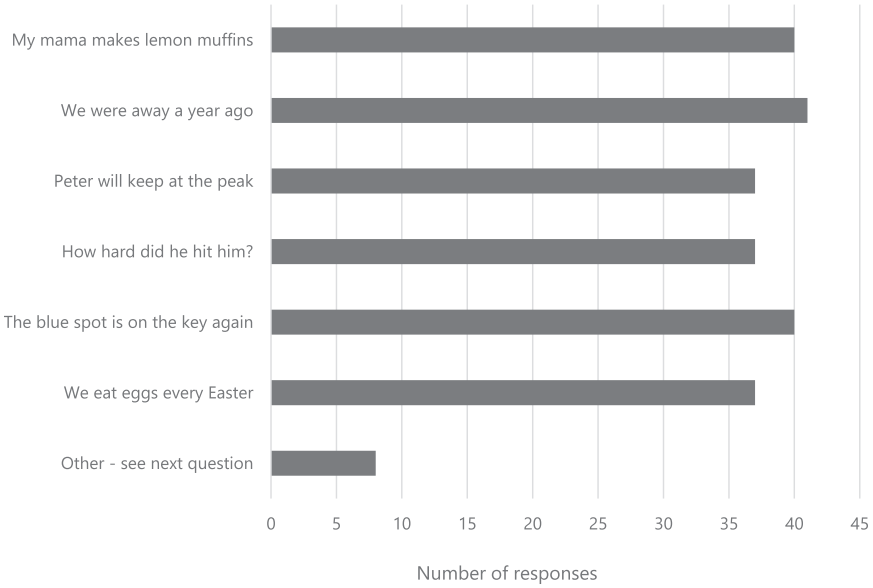

FIGURE 2. Frequency counts for CAPE-V sentences elicited by participants in a typical administration of the CAPE-V protocol (Question 11).

they changed four of the six sentences used on the form, with the following rationale:

> "We changed these sentences to remove a religious reference (Easter), remove the violence reference (hitting), and make it more sensical (we noticed patients stumbling over "Peter will keep at the peak"). The "lemon jam" is listed in the original CAPE-V 2009 article."[b]

Spontaneous speech prompts reported by 43 of the participants are in Table Q13 in Supplementary Materials. Briefly, 14 reported using the prompt "Tell me about your voice" or something very similar. Some wrote that they obtain spontaneous speech samples as they conduct the interview or informally chat with the patient. Others ask the patient to talk about their lives, favorite things, or happy memories. One simply asks the patient to speak for 30 seconds.

Although there is no reading passage in the CAPE-V protocol, we were aware that some clinicians obtain reading samples as part of their assessment of voice quality, so we included it as an option in Question 14. All 24 participants who responded to the question mentioned the Rainbow Passage, with one also using the Grandfather Passage[7] (Darley et al) and one also using the Caterpillar Passage (Patel et al).[8]

As displayed in Figure 3, 76% of respondents reported spending less than 5 minutes eliciting CAPE-V stimuli.

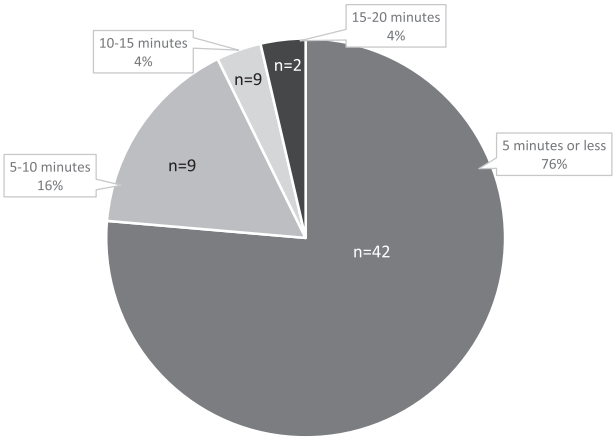

FIGURE 3. Reported time to elicit CAPE-V stimuli (Question 15).

Rating tasks

Question 16 asked who typically provides voice quality ratings, and allowed respondents to choose more than one answer. Fifty four chose SLP, 20 chose students, and one chose ENT. Two reported a mix of either SLP/ENT or SLP/student providing voice quality ratings.

In response to Question 17, 52 of the 54 participants who responded reported rating voice quality during or immediately after the evaluation session. Despite this, 25% of the 44 respondents to Question 18 reported that they "always" or "usually" base their ratings on a recording, with 45% writing that they "rarely" or "never" used the recording to rate voice quality. Of those who report basing their ratings on recordings (n = 34), 43% reported they listen only once before making a decision (Table Q19 in Supplementary Materials).

All of the 49 respondents to Question 20, which asked about the voice parameters used, reported rating Overall Severity, and most rate the other CAPE-V parameters at least sometimes (Figure 4). Most participants who responded to Question 21 reported that they also rate resonance (n = 43), instability (n = 32), or some "other" voice parameter (n = 21; Tables Q21a and b in Supplementary Materials).

[b] NB: "Lemon jam" is used in the article and instructions in Appendix B of Kempster et al[1]; however, "lemon muffins" is on the CAPE-V form published therewith.

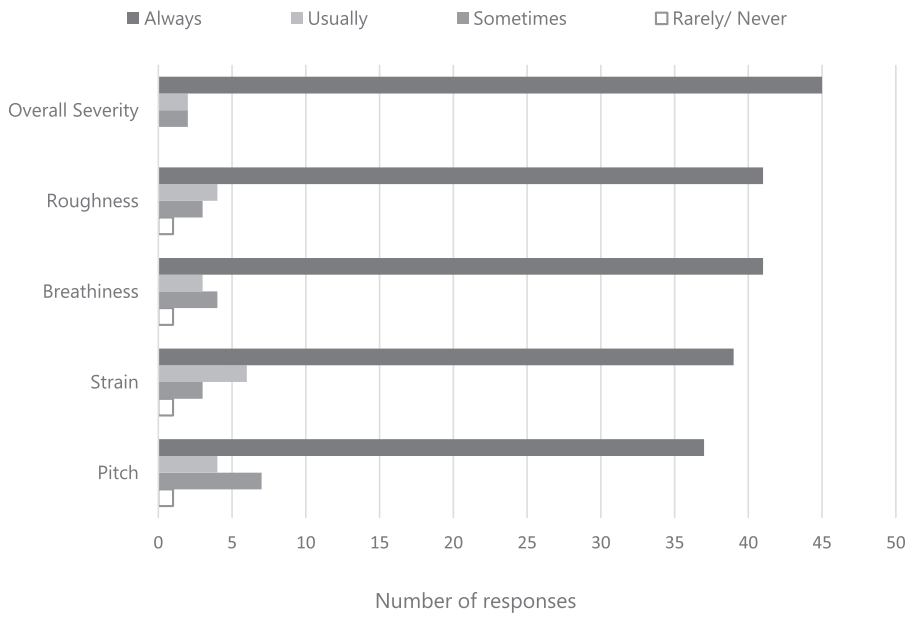

FIGURE 4. Frequency counts for participants' rating of the main voice quality parameters in the CAPE-V protocol (Question 20).

Fewer than half of the 49 respondents to Question 22 reported that they routinely or sometimes use the blank scales on the CAPE-V form (Tables Q22a and b in Supplementary Materials). The four participants who do not use the blank scales reported writing comments in their notes, assigning a number based on knowing where the categorical markers are, and similarly, "converting to a numerical scale" using eight categories of severity (eg, minimal = 1-9, mild = 10-20, mild-moderate = 21-35).

Consistency. Of the 49 participants who responded to Question 23, 39% stated that they always note the consistency of all or some voice parameters. One other reported "I almost always rate consistency, but find consistent or intermittent is not sufficient. I often break intermittent into intermittent – frequent or intermittent – sparse." More than half reported that they only note consistency of voice quality if it is noticeably intermittent, for example, "Often, for roughness. Other qualities such as loudness tend to be more consistent." Two percent rarely or never make a note of the consistency of any voice parameter.

About half of the 48 participants responding to Question 24 reported noting consistency/intermittency by circling the letters I or C on the CAPE-V form; about 38% make text notes on the form. Seventeen percent reported "other," which included some providing more detail about the context of consistency (ie, "end of sample vs start of sample" Tables Q24a and b in Supplementary Materials). Several participants responded that they do not use the form at all.

Finally, 82% of participants reported rating and scoring the CAPE-V in less than 5 minutes, as shown in Figure 5. None reported taking longer than 15 minutes.

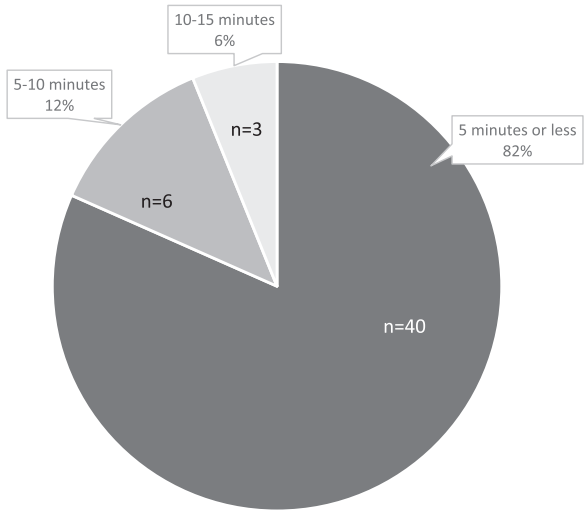

FIGURE 5. Reported time to complete rating and scoring CAPE-V results (Question 25).

Open-ended questions

Skipping sections. Sixty-one percent (n = 29) of the participants who responded to Question 26 indicated that they do not skip any sections of the CAPE-V. The reasons for skipping sections provided by the remaining 39% are shown in Table Q26 in Supplementary Materials. These reasons were related to the efficiency and functionality of CAPE-V tasks, and most of them could be summarized by one or more of three broad themes.

Theme 1. The scoring protocol is too detailed. Many responses to the question of skipping a task or section of the CAPE-V protocol indicated that users find the protocol or form too complex. This has mainly to do with the number of parameters, and consequently the number of scales, to be completed as part of the protocol. One participant indicated that they rate Overall Severity using the scale, but for other parameters, they comment in text. Another wrote "I use mild, moderate, severe as my markers and not a

numerical value," suggesting that it is more functional for them to categorize and describe voice quality in words rather than to issue a rating. One participant addressed the challenge of assigning a number to a ratio scale when thinking in terms of intervals, writing "I don't use the visual analog scale (VAS) scales and just write a number in levels of 5 (5, 10, 15, 20, etc) because I do not have 100 different levels of perception for a specific auditory-perceptual construct." This participant apparently recognizes that scoring in intervals of 5 reduces the continuous VAS into 20 segments (hence, a 21-point ordinal scale) and even considers less precision, adding "I've been toying with just doing it in 10's, because the 5's seem a little too much for me."

Theme 2. Doing the entire protocol is too time-consuming. Although the majority of participants in this study reported taking less than 10 minutes to do elicitation, rating, and scoring of CAPE-V tasks, the most commonly reported reason for skipping tasks was that the entire protocol takes too much time. Some of this was clearly driven by the paper-based modality of typical CAPE-V use, as stated clearly by one participant: "I… haven't got time to pull out a ruler and measure my "x" marks." Other clinicians reported emphasizing different parts of voice evaluation to the detriment of auditory-perceptual evaluation, with one writing "I complete other patient-reported outcome measures and more objective ratings first, so if I don't have time, I will skip it altogether."

Theme 3: Stimuli or tasks are irrelevant or overlap with other components of assessment. Some clinicians had strong feelings about the utility of some of the CAPE-V tasks. One participant admitted "We only collect the vowel /a/, because we haven't noticed a difference between /a/ and /i/, although that is an anecdotal observation, not supported by any study." Another indicated that they find the sentence stimuli to be irrelevant and are not convinced they serve a "functionally sound purpose," despite the protocol's physiophonetic explanations for including both vowels and all of the sentences in the protocol. Relatedly, one participant described a disconnect between the relevance of voice quality judgments based on the read sentences compared with "everyday" voice quality, writing "I skip sentences and reading because spontaneous speech is what I want to change in therapy, so I 'cut the fat' and only go with what I'm interested in changing." This seems to be the position of another participant, who reported choosing only the stimuli that seem relevant for a given patient: "The most important thing to me is what do they sound like when they are speaking in spontaneous speech. If I have the need to look at something specific, for example, laryngeal dystonia testing, I'll do specific tasks as indicated."

On the whole, participants recognized the duplication of some tasks from the CAPE-V within a typical voice evaluation. They seemed to choose not to complete segments of the CAPE-V protocol or form that they believe they might otherwise score 0 or "within normal limits." Comments related to functionality revealed some strong opinions about the content of the sentence stimuli and the visual analog scales used to rate voice parameters. Two participants also addressed the fact that they use the CAPE-V with populations for whom the current tasks or stimuli are inappropriate (ie, children, altered mental status).

Suggested changes. Half of the 46 participants responding to Question 27 had no recommendations for improving the CAPE-V, but the other 23 provided at least one (see Table Q27 in Supplementary Materials). Most of these suggestions were associated with three broad themes, involving revising or removing tasks; clarifying terminology and operational definitions of the some of the parameters; and simplifying or automating the rating procedures.

Theme 4: Remove or modify tasks. Many of the suggestions shown in Table Q27 were directed at modifying or removing some of the existing elicitation tasks, although one participant did suggest adding a counting task. Many seemed to believe erroneously that a longer reading task (ie, The Rainbow Passage) is part of the protocol, and in general, participants appeared to crave the flexibility to skip some or all of the sentences. To put it bluntly, as one participant did: "Just do spontaneous speech and take out the sentences and reading passage."

Relatedly, and as mentioned in responses to Question 26, participants also raised doubts about the relevance of the tasks, suggesting they are unrepresentative of their patients' daily voice use. These participants seemed to be objecting to the whole idea of a standard set of stimuli (eg, "I don't believe that a standard set of stimuli is functionally relevant to the patient and results in different vocal behavior than they use in their daily lives.").

A few participants seemed to indicate that they would prefer to follow the protocol, but that it did not accommodate some of the populations they serve. One participant suggested creation of a pediatric version of the CAPE-V or that it be separately validated for children. No specific modifications were proposed, but it seems likely that sentences are the most likely challenge for the pediatric population. Another participant proposed "Mak(ing) it easier to obtain with altered mental status patients." Although it is not clear which changes would simplify the protocol for patients with dementia or psychosis, this clinician may be seeking guidance on how to administer the CAPE-V based on limited patient cooperation.

Finally, some specific modifications to the CAPE-V sentences were suggested. Many comments were directed at their cultural insensitivity and awkwardness (eg, "How hard did he hit him?"). Numerous participants reported that they have already modified the sentences in their own practice to make them "more culturally sensitive… and trauma informed," because the sentences are "a little dated and nonsensical" (see Table Q12 in Supplementary Materials). The semantic unpredictability of some of the sentences leads many patients to "incorrectly read 'Peter will keep at the peak' as 'Peter will keep at the park'," for example.

Theme 5: Clarify the terminology and parameters. There was a desire for clarity on the meaning of some of the CAPE-V terms; for example, one participant wrote "I feel strain is the most difficult to define and identify…." Many comments clearly requested frank explanations of concepts named in the CAPE-V protocol, but some were not fully interpretable without more context. For example, one participant asked for "specific outlined parameters for rating severity," and to "make it less subjective on severity rating." The latter could be interpreted as an appeal for a better understanding of what "overall severity" means as a parameter, or as a request for clarification on rating the severity of other parameters. Another participant reported "I think pitch could be reframed to be more relevant." This may be a reference to the non-committal nature of the pitch scale on the form, which requires that the user write in the direction of deviance; on the other hand, it could be a request to justify the role of pitch ratings in the presence of standard acoustic measures of habitual fundamental frequency.

Theme 6: Simplify or automate rating procedures. Consistent with reports of evaluating voice quality based on mental or "guesstimated" ratings, there were some requests to simplify the rating procedure itself or to reduce it to an ordinal scale. For example, one participant suggested that we "Remove VAS and make it an estimate in 10's or categories (mild, mild-mod, mod, mod-severe, severe)."

There were also requests for a simplified tool for measurement, for example, "A tool that is a ruler that you can overlay the scale, with clear indicators of mild versus mod versus severe boundaries." Relatedly, although the iCAPE-V is available to clinicians with access to Pentax's Computerized Speech Lab, one participant requested "an electronic version with sliders that generated a quick text report with numbers."

Finally, one participant described an efficient way of contemporaneously listening and rating: "Honestly, I 'rate' the perceptual part while I'm also taking my measures (CPP, Range, Mean Fo, etc) so I follow my laryngeal function studies protocol more than the actual CAPE-V protocol."

Requests for follow-up. Only one respondent asked to be contacted by the first author to further discuss the CAPE-V. This was an experienced CAPE-V user and researcher who said in a virtual interview that when they started thinking about the CAPE-V, "I realized I came up with my own rules." Their comments are consistent with Themes 1 and 6, described above, and are attached as Table Q28 in Supplementary Materials.[c] This participant described making auditory-perceptual judgments not only during evaluation sessions, but also at the end of each treatment session, rating only Overall Severity, based on spontaneous speech, presumably in multiples of 5.

[c] NB: This participant had already contributed to the survey results; to avoid overrepresenting their opinions, only novel comments from the interview are included here.

DISCUSSION

This survey provided end-user feedback on the clinical use of the CAPE-V and suggestions for improving the instrument in the future. Most of the 49 participants who completed the survey were attendees at the 2023 Fall Voice Conference in Washington, DC; others were members of the associated discussion forum, "Fall Voice Community: Eat, Breathe, Talk." All of the sources of variability identified on our conference poster were confirmed to varying degrees by the sample of clinicians who participated in this study.

Eliciting stimuli

As predicted, some clinicians reported recording stimuli; others rate productions only in real time. The majority of clinicians, who may have switched to virtual evaluations in the first few years since the onset of the COVID-19 pandemic, have since resumed performing voice evaluations in-person. We were surprised to find, however, that only 40% of those who reported doing in-person evaluations either audio- or video-record components of the evaluation. An even higher proportion of those who conduct virtual evaluations record nothing. Because of the benefits of listening to the voice samples multiple times, rating voice quality after the clinical visit, and of retaining pretreatment recordings as a baseline reference, the CAPE-V instructions specifically require audio recording of elicited stimuli, even providing specific information about appropriate mouth-to-microphone distance, resolution, and sampling rate. Two-thirds of participants reported rating voice quality in the presence of the patient (ie, during the evaluation itself). Those who do record the evaluation tend to listen to the stimuli once or twice before rating; only six of the 34 who responded to Q19 listen more than two times, but 5 of these wrote that they listen "as many times as needed." It should be noted that the CAPE-V instructions say nothing about repeated listening, neither prohibiting, limiting, nor recommending it. But, given that clinicians who do not make recordings will not even have the option of repeated listening, the basis on which ratings are made appears to be another major source of variability. Solomon et al[9] compared ratings of voice quality from a live clinical situation and in a controlled laboratory situation that included auditory anchors and excluded clinical history, and found that clinical ratings of Overall Severity, Roughness, Breathiness, and Strain were higher (more severe) than laboratory ratings. Although the laboratory methods were more austere than those that would be used in a typical clinical environment, this work provides some evidence that ratings from live versus recorded stimuli are likely to differ.

We expected to find that participants modified or skipped some of the CAPE-V tasks and we were interested in their rationale. As anticipated, participants reported a variety of discrepancies in their implementation of the protocol, including eliciting different sentences, fewer vowels, and basing their voice quality ratings on other components of the evaluation, such as the initial interview, production of pitch glides, and reading a passage of text. As shown in Table Q26 in Supplementary Materials, although many were motivated by

time constraints, more than a few disputed the necessity or even the validity of the components of the CAPE-V protocol or form. There seem to be complementary issues here: The rationale behind including parts of the protocol is not clear to some users, and without a clear reason to perform a task, clinicians are uncomfortable spending time doing it.

It appears that many clinicians may benefit from reconsidering their methods within the parameters of their work settings. Specifically, and as noted in the original CAPE-V documentation, there is overlap between the CAPE-V tasks and the stimuli routinely produced during laryngoscopy (ie, vowels) and the initial interview (ie, spontaneous speech). The only tasks that are specific to the CAPE-V are the sentences. We suspect that those who report eliciting CAPE-V stimuli in 5 minutes or less are already capitalizing on this overlap, and that others may benefit from rearranging their approach to eliciting relevant stimuli.

Rating voice quality

One issue of concern that we expected to arise from the survey was the presence of two versions of the CAPE-V scales. Specifically, versions of the CAPE-V form released by ASHA Special Interest Division 3 showed 100-mm hybrid VA scales with nonlinearly spaced text markers (MI/MO/SE) centered at 10, 35, and 72 mm.[10,11] The version published in the peer-reviewed paper describing the rationale and development of the CAPE-V showed equally spaced text markers at 10, 50, and 90 mm.[1] Previous work has shown that the presence and location of subscale text markers affects ratings of voice quality, particularly for women's voices.[12] The current study did not inquire about which version of the CAPE-V scales the clinicians use, but we established that many of them used idiosyncratic scoring methods, including "guesstimating" the position of their mark on a visible or mental VAS; reporting ordinal ratings (eg, multiples of 5 or 10), or simply stating categories of severity (eg, mild-moderate-severe), the meanings of which might differ markedly depending on which scale was used. All of these practices increase the variability and inconsistency of CAPE-V administration.

Because only half of the participants surveyed by Nagle[5] circled the "C" or "I" on the CAPE-V form, we were interested in how participants noted consistency or intermittency in any of the voice parameters. Most of the clinicians in the current study reported noting at least the intermittency of those voice characteristics that are not consistent, and of those, nearly half said they circle the "C" or "I." Others reported that they make a note on the form, but three participants revealed in this response that they do not use the form at all. This raises the serious question of whether they should be claiming to administer the CAPE-V at all, and if so, whether are they executing the tasks from memory or rating the parameters systematically.

Fifteen percent of the participants in Nagle's 2022 study of experienced voice clinicians admitted rating only Overall Severity, reporting that they always skip the dimensions of Roughness, Breathiness, and Strain. Accordingly, we wondered which voice parameters the current cohort of survey participants routinely rate. As shown in Figure 4, a large majority reported rating all six parameters listed on the form. A small number of participants seem to rate only Overall Severity, but some of them reported doing so only sometimes. This was surprising to us, because rating Overall Severity seems to be the minimum one can do, again raising the question of what clinicians mean when they say they "did a CAPE-V." Anecdotally, we have found that "doing a CAPE-V" can refer to completing the entire protocol as directed; eliciting only the CAPE-V sentences; using the only the CAPE-V form; or just using a scale of some kind to rate voice quality based on some stimulus.

Unlike the previous scales on the CAPE-V form, the Pitch and Loudness scales require the user to "Indicate the nature of the abnormality," providing a short line to do so. We did not probe how our participants indicate the direction of deviance in pitch or loudness; it could be that they mark the scale, make notes on the form, or some combination of both. However, given that acoustic correlates of pitch and loudness should be obtained as part of a standard voice evaluation, it was frankly surprising to find out that 96% of respondents rate both parameters at least some of the time. This indicates a recognition by these clinicians that measures of sound pressure level and voice fundamental frequency are imperfect correlates for auditory-perceptual judgments of loudness and pitch. On the other hand, it could reveal a decision to skip acoustic measures altogether in favor of providing a quick perceptual rating.

Suggestions for changes to the CAPE-V

Based on the quantitative results and thematic findings of this study, few, if any, users of the CAPE-V administer the protocol strictly according to the instructions. Their reasons for modifying the protocol and form seem to be driven mostly by an overall desire for efficient use of limited evaluation time with patients, as reflected in five of the six themes identified. Some of the comments seem to be based on the unwieldiness of paper forms and rating scales (and the requirement of the protocol to measure marks on the VAS with a ruler); others indicate a decision to eliminate elicitation tasks, not always with a clear rationale. Regardless of their rationale, users are cautioned that varying from the standard CAPE-V protocol violates the spirit of "doing a CAPE-V" and can undermine the integrity of the instrument.

It is a concern that only about half of the respondents to this survey reported that they record voice productions during evaluation, particularly as most clinicians ostensibly obtain acoustic and aerodynamic measures using similar stimuli. This suggests at least three possibilities: (1) some are prevented from doing so because of logistical or time constraints; (2) some do not anticipate comparing pre-/post-treatment data (for whatever reason); and (3) some do not understand the relevance of collecting and maintaining a historical record of patients' voice quality.

As described in Theme 5, there is an apparent desire among many clinicians for better consensus on the terminology and concepts used in the CAPE-V protocol,

including the most general parameter of Overall Severity. Part of addressing this need is identifying a group of stimuli that might be chosen to exemplify mild, moderate, and severe degrees of each of the CAPE-V parameters, paired with a set of ratings and descriptions of the stimuli obtained from experienced voice clinicians, as described by Nagle.[5]

Finally, some of the responses to the open-ended questions revealed a lack of knowledge or interest in the CAPE-V authors' rationale for including and excluding what they did. One particularly blunt comment captures this: "When I see the CAPE-V, I imagine a lot of academics who don't have to see people in clinic all day (no offense…) or a bunch of grad students who need a high level of structure because they don't know what they're listening for yet." Relatedly, more than one clinician noted that they were interested in eliciting only everyday voice productions (ie, extemporaneous speech) during an evaluation, for example, one wrote "I find the specific stimuli to be irrelevant and I am not convinced they serve a functionally sound purpose." Feedback such as this indicates a need to justify more clearly the reasoning behind many aspects of the CAPE-V protocol and form.

Limitations

Because the survey was anonymous and shared via a link with participants who may not have attended the Fall Voice Conference, we obtained very little demographic data, which limits our ability to determine the amount of bias in our sample. Clinicians who attend the Fall Voice Conference are likely to work in settings with a relatively high proportion of voice patients, and although many of them may not use the CAPE-V, they are also likely to have a greater-than-usual interest in voice evaluation in general and auditory-perceptual parameters of voice in particular.

Although 88 individuals opened the survey, only 59 proceeded past the consent form and only 49 completed the survey through Question 25. We cannot know why some respondents did not continue the survey, although it is possible that they were curious non-CAPE-V users or felt that they lacked the experience with the CAPE-V to answer questions about it without a reference to the form itself. Our sample was not large enough to examine the associations among demographic characteristics and clinical practices.

Future research

Given the variability of implementation of the CAPE-V reported here and elsewhere,[4,5] we have begun to examine what clinicians typically do in routine voice evaluations, particularly when working as part of a multidisciplinary team. For example, from conversations during data collection for this study with clinicians who do not use the CAPE-V, we learned that some clinicians, and many ENTs, continue to use the GRBAS scales,[13] frequently in conjunction with the CAPE-V sentences. Additional studies are needed to establish the prevalence of GRBAS usage in relation to the CAPE-V, at least among SLPs.

Although the majority of participants reported that it takes them 10 minutes or less to complete both administration and scoring of the CAPE-V, one of the most-mentioned reasons for skipping CAPE-V tasks was that they were too time-consuming. Although few participants reported using the iCAPE-V or other electronic versions of the form, it seems likely that an electronic version of the protocol and form could address some of the participants' time concerns.[4]

The evident desire of many clinicians for "permission" to add, remove, or modify aspects of the CAPE-V protocol makes it clear that a revision of the CAPE-V protocol is in order. Ideally, this would involve continued input from experienced users and a greater emphasis on the rationale for every part of the CAPE-V protocol. It has become clear that maximizing fidelity to a revised CAPE-V protocol will require an implementation strategy that includes consideration of the logistical constraints on clinicians and the fact that many do not evaluate voice quality on a regular basis. Freely available training programs (eg, All-Voiced[14] and the University of Wisconsin Voice Simulations and Games website[15]) are a critical component of this strategy, guiding unfamiliar users through the process of administering and scoring the CAPE-V, and providing access to a set of recordings accompanied by expert ratings against which clinicians can compare their scores.[16] In addition, research is needed to demonstrate the efficacy of training programs.[17] Relatedly, for those who truly cannot elicit, record, rate, and score two vowels, six sentences, and a sample of spontaneous speech in their allotted time with a patient, an abbreviated version of the CAPE-V may be feasible, provided that research confirms that it would yield valid results.

CONCLUSION

The results of this survey provide further evidence that voice-specialized clinicians do not faithfully follow the instructions for the administration of the CAPE-V, information that we have gathered over previous surveys, and interviews.[4,5] In this exploratory study, queries elicited from a larger group of voice clinicians focused on specific tasks, measurement approaches, reasons for skipping items, and suggestions for revision. From the latter two topics, we determined six themes that should be considered in a revision of the CAPE-V. Such revisions are currently in progress, and involve modifications to the form, tasks, stimuli, and rating procedures.

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

Data availability

All data generated or analyzed during this study are included in this published article and its supplemental material files.

Declaration of Competing Interest

The authors have no relevant financial interests.

APPENDIX A: FALL VOICE SURVEY 2023

This is an anonymous survey conducted by Kathleen Nagle, PhD, CCC-SLP, investigating how clinicians use the CAPE-V in their daily practice. It has been approved by the Seton Hall Institutional Review Board (#2024-493). Your consent is implied by continuing with the survey. If you do not wish to participate in this study, please exit the browser.

Q1 For how long have you been using the CAPE-V to evaluate voice quality?

○ One year or less (1)
○ 1-5 years (2)
○ 5-10 years (3)
○ More than 10 years (4)

Q2 For how many years has voice evaluation been part of your practice? Please round up to an integer.

_____

Q3 What proportion (%) of your current practice is voice evaluation?

_____

Q4 In what type of setting do you practice? Choose all that apply for voice evaluation.

☐ Private clinic (1)
☐ Hospital clinic (2)
☐ University clinic (3)
☐ Other (please explain) (4)

Q5 How many voice evaluations do you typically do in a week?

_____

Q6 In what environment(s) do you evaluate your voice patients?

○ All in-person (1)
○ Mostly in-person (2)
○ About 50/50—in-person/virtual (3)
○ Mostly virtual (4)
○ All virtual (5)
○ Other (please provide details) (6)

Q7 When evaluating a patient in-person, how do you typically administer the CAPE-V?

☐ On paper (1)
☐ Electronically, clinic-developed program (2)
☐ Electronically, web-based program (ie, Pentax program) (3)
☐ Other (please explain) (4)

Q8 When evaluating a patient in-person, do you record their voice?

○ Yes, with audio and video (face tape) (1)
○ Yes, with audio only (2)
○ No (3)
○ Other (please explain) (4)

Q9 When evaluating a patient virtually, do you record their voice?

○ Yes, audio and video (1)
○ Yes, audio only (2)
○ No (3)
○ Other (please explain) (4)

Q10 What stimuli do you elicit from patients when you administer the CAPE-V? Please choose all that apply.

☐ CAPE-V sentences (1)
☐ Vowel /a/ (2)
☐ Vowel /i/ (3)
☐ Spontaneous speech (4)

☐ Reading passage (5)
☐ Other (please explain) (6)

Q11 Which CAPE-V sentences do you use?

☐ My mama makes lemon muffins (1)
☐ We were away a year ago (2)
☐ Peter will keep at the peak (3)
☐ How hard did he hit him? (4)
☐ The blue spot is on the key again (5)
☐ We eat eggs every Easter (6)
☐ Other—see the next question (7)

Q12 What other sentence stimuli do you use? Please explain, if you like.

_____

Q13 What prompt(s) do you use to elicit spontaneous speech?

_____

Q14 What reading passage(s) do you use?

_____

Q15 How long does it take you to *elicit* the CAPE-V stimuli?

○ 5 minutes or less (1)
○ 5-10 minutes (2)
○ 10-15 minutes (3)
○ 15-20 minutes (4)
○ Usually longer than 20 minutes (5)

Q16 Who typically rates the patient's voice quality specifically using the CAPE-V at your work site? Please select all that apply.

☐ SLP (1)
☐ SLP student (2)
☐ ENT (3)
☐ Other (please explain) (4)

Q17 When do you typically rate a patient's voice quality?

○ While administering the CAPE-V, with the patient present (1)
○ Immediately after the session (2)
○ Some time after the session (3)

Q18 How often do you use recordings of the voice to rate voice quality (as opposed to rating in real time or relying on memory)?

○ Always (4)
○ Usually (3)
○ Sometimes (6)
○ Rarely or never (1)
○ Other (please explain) (5)

Q19 When you base your ratings of voice quality on recordings, how many times do you listen to each stimulus?

○ Just once (1)
○ Twice (one repetition) (2)
○ More than twice (3)
○ As many times as needed (4)

Q20 What voice quality parameters do you rate when using the CAPE-V, and how often do you rate them? (If "Other," please note on the next question)

	Always (1)	Usually (2)	Sometimes (3)	Rarely or Never (4)
Overall Severity (1)	○	○	○	○
Roughness (2)	○	○	○	○
Breathiness (3)	○	○	○	○
Strain (4)	○	○	○	○
Pitch (5)	○	○	○	○
Loudness (6)	○	○	○	○

Q21 What characteristics of voice quality other than the above do you usually note?

☐ Resonance (1)
☐ Instability (2)
☐ Other (please explain) (3)

Q22 Do you use the empty scales on the CAPE-V form to rate the extra voice characteristics?

○ Yes (1)
○ Sometimes (please explain) (2)
○ No (3)
○ Other (please explain) (4)

Q23 How often do you note consistency/inconsistency of a voice quality parameter?

○ Always, for every parameter (1)
○ Always, but only for certain parameters (please explain) (2)
○ Only if it is inconsistent/intermittent (3)
○ Rarely or never (4)
○ Other (please explain) (5)

Q24 How do you indicate that voice quality is inconsistent or intermittent?

☐ Circling or choosing "I" on the form (1)
☐ Writing inconsistent or intermittent on the form (2)
☐ Other (please explain) (3)

Q25 How long does it take you to *rate and score* the CAPE-V?

○ 5 minutes or less (1)
○ 5-10 minutes (2)
○ 10-15 minutes (3)
○ 15-20 minutes (4)
○ More than 20 minutes (5)

Q26 If you skip sections of the CAPE-V procedure or stimuli, what is your rationale? If you do not skip any of the CAPE-V sections, please write N/A.

_____

Q27 If you could change the CAPE-V, what would you add, revise, or remove from the protocol or form? If you would not change a thing, please write N/A.

_____

Q28 If you have more to say about the CAPE-V and would like to be contacted by the PI, please add your email here. [If you have questions, concerns, or complaints about this research project, you can contact the Seton Hall University Institutional Review Board (IRB) at (973) 761-9334 or irb@shu.edu. You may contact the Principal Investigator at naglekat@shu.edu.]

_____

APPENDIX B: FALL VOICE POSTER 2023

Toward Recommendations for a New and Improved CAPE-V

Kathleen F. Nagle[1], Gail B. Kempster[2] & Nancy P. Solomon[3]

[1]School of Health & Medical Sciences, Seton Hall University, Nutley, NJ
[2]College of Health Sciences, Rush University, Chicago, IL
[3]Walter Reed National Military Medical Center, Bethesda, MD

2023 Fall Voice Conference
Washington, DC

Rationale

The **Consensus Auditory Perceptual Evaluation of Voice (CAPE-V**; ASHA, 2002; Kempster et al., 2009) was developed to standardize the evaluation of voice. Variability in how clinicians follow the protocol and use the form to rate voice quality (fidelity) has prevented development of training protocols that might improve agreement among clinicians. Some of the main sources of variability are:

- There are **two versions of the CAPE-V form**, with different labeling of the rating scales (Nagle et al., 2014)
- Some stimuli are **modified or skipped** by some clinicians (Zraick et al, 2011)
- Not all clinicians **mark consistency** within or across stimuli (Lodhavia & Kempster, *under review*; Nagle, 2022)
- Some clinicians **record stimuli**; others only rate them in real time (Lodhavia & Kempster, *under review*)
- Some clinicians rate **only overall severity** and not the other main parameters (Nagle, 2022)

Tell us how **you** use the CAPE-V at our 10-minute **anonymous survey**

Discussion

- Neither extant version of the CAPE-V form (Fig.1a, 1b) is psychometrically ideal; both have textual labels that affect the validity of each scale line as a visual-analog scale (VAS). A **traditional VAS** (i.e., line with labels only at the endpoints) would be more appropriate.
- The CAPE-V protocol would benefit from **revisions to the sentence stimuli**. If the given stimuli are truly meant to "elicit different laryngeal positions," the sentences can be revised to be more phonetically and socio-culturally appropriate.
- Consistency markers may serve as a reminder to some.
- What does it mean to rate voice quality in real time, compared to rating based on an audio recording? Do results differ if voices are rated based on **recorded samples vs real-time live productions**?
- Voice quality is multidimensional, but it can be difficult to isolate aspects of breathiness, roughness and strain for separate ratings. Can a dysphonic voice be **adequately characterized with a single rating of overall severity,** along with a narrative description?

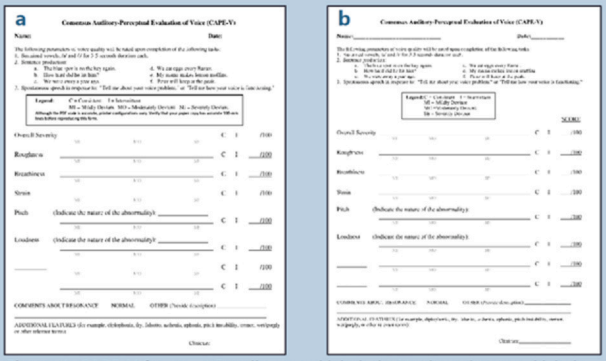

Figure 1. CAPE-V forms. a) Equally spaced labels; b) Nonlinearly spaced labels. Reprinted with permission from ASHA (2002; 2009); copyright 2009 ASHA.

- Based on this survey, interviews with clinicians and personal experience, this team of authors is developing an **updated version of the CAPE-V**.

References

ASHA. Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) ASHA Special Interest Division 3, Voice and Voice Disorders. Published online 2002. Accessed August 4, 2009.
Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. American Journal of Speech-Language Pathology. 2009;18(2):124-132.
Lodhavia A, Kempster GB. Fidelity to the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). Submitted for publication 2023.
Nagle KF. Clinical use of the CAPE-V scales: Agreement, reliability and notes on voice quality. Journal of Voice. 2022; online.
Nagle KF, Helou LB, Solomon NP, Eadie TL. Does the presence or location of graphic markers affect untrained listeners' ratings of severity of dysphonia? Journal of Voice. 2014;28(4):469-475.
Zraick RI, Kempster GB, Connor NP, et al. Establishing validity of the consensus auditory-perceptual evaluation of voice (CAPE-V). American Journal of Speech-Language Pathology. 2011;20(1):14-22.

Appendix C. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jvoice.2024.08.032.

References

1. Kempster GB, Gerratt BR, Verdolini Abbott K, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol.* 2009;18:124–132.
2. 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.
3. 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.
4. 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.
5. 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.
6. Lingard L. Beyond the default colon: effective use of quotes in qualitative research. *Perspect Med Educ.* 2019;8:360–364. https://doi.org/10.1007/s40037-019-00550-7.
7. Darley FL, Aronson AE, Brown JR. *Motor Speech Disorders.* Philadelphia, PA: Saunders; 1975.
8. Patel R, Connaghan K, Franco D, et al. "The caterpillar": a novel reading passage for assessment of motor speech disorders. *Am J Speech Lang Pathol.* 2013;22:1–9. https://doi.org/10.1044/1058-0360(2012/11-0134).
9. 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.
10. 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.
11. 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.
12. Nagle KF, Helou LB, Solomon NP, et al. 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.
13. Hirano M. "GRBAS" scale for evaluating the hoarse voice & frequency range of phonation. In: Hirano M, ed. Wien: Springer-Verlag; 1981:88–89. Clinical Examination of Voice. Vol 5. Disorders of Human Communication.
14. Calaf N. All-Voiced [Web application]; 2024. Available at: https://www.all-voiced.com/. Accessed August 20, 2024.
15. Connor NP, Bless D, Dardis C, et al. DoIT Engage Project Team. Simulations: Consensus Auditory-Perceptual Evaluation of Voice [CAPE-V]. Voice Disorders: Simulations & Games; 2008. Available at: https://csd.wisc.edu/slpgames/sims.html. Accessed August 20, 2024.
16. Walden P. Perceptual Voice Qualities Database (PVQD). Published online October 9, 2020. Available at: https://data.mendeley.com/datasets/9dz247gnyb/3. Accessed October 20, 2020.
17. Franca MC, Boyer VE, Tripathee P. Effects of training in voice auditory-perceptual skills. *Int J Lang Commun Disord.* 2024:1–12. https://doi.org/10.1111/1460–6984.13103.