Fidelity to the Consensus Auditory-Perceptual Analysis of Voice (CAPE-V): A Pilot Study

[*,†]Anjli Lodhavia, and [*,‡]Gail B. Kempster, *‡*Chicago, Illinois, and* †*West Bloomfield, Michigan*

Summary: Objectives. Auditory-perceptual measurements of voice are among the most common diagnostic tools used during a voice evaluation and are considered a gold standard for documenting voice disorders. The goal of this pilot study was to examine the fidelity of clinicians in the administration of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) and its published protocol. This investigation aimed to determine how the CAPE-V is being used by clinicians and researchers and the extent to which users deviate from the published procedure.

Study design. Survey

Methods. Data were collected by surveying a group (N = 17) of speech-language pathologists who regularly evaluate and treat patients with voice disorders. Survey results revealed that few of the sampled speech-language pathologists follow exactly the standardized instructions for administering the CAPE-V. Considerable variability in CAPE-V administration, including tasks and stimuli examined, was found across respondents.

Conclusions. This exploratory project may be used to develop a larger national survey study investigating fidelity to the CAPE-V and to motivate recommendations for possible revisions to the instrument and its protocol for administration.

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

INTRODUCTION

The Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) is a clinical measurement tool that was developed to promote a standardized approach to auditory-perceptual measurement of vocal quality, including pitch and loudness. The original protocol for the CAPEV was first made available by ASHA in 2002 and was published in a peer-reviewed journal with slight modifications in 2009.[1] The original instrument's validity was supported in a follow-up study.[2] Since its inception, the CAPE-V has been used in clinical and research settings around the globe. The 2009 publication describing the development of this tool and directions for its implementation earned a distinction as being one of the most frequently cited publications in laryngology since the year 2000.[3] Nonetheless, it is unclear how closely those who use this evaluation tool adhere to the specified protocol, indicating procedural fidelity to the tasks and procedures provided by the original authors.

Fidelity is the extent to which the administration of an evaluation or treatment tool matches the prototypical gold standard that was first introduced empirically.[4] Kaderavek and Justice assert that fidelity is a construct that is clinically relevant to the field of speech-language pathology, particularly to increase compliance with evidencebased practice (EBP) standards.[5] Utilizing EBP with patients ensures that all individuals are assessed and treated using scientifically-based tools. Because of the emphasis on EBP across healthcare disciplines, fidelity is critical for documenting the effectiveness of both assessment and intervention methods in all areas of practice, including the evaluation of voice disorders.

Fidelity to an assessment or treatment tool is likely to lessen when speech-language pathologists (SLPs) implement EBP standards in the clinical setting. Since variation effects are inevitable, achieved fidelity will never match intended fidelity. However, all detailed procedures for an assessment or intervention method must be consistently applied to produce the best outcomes in clinical practice.[5] Low procedural fidelity to an assessment tool, such as the CAPE-V, can negatively affect the communication and transferability of diagnostic findings, as well as affect the clinical decision-making process for appropriate intervention approaches. Therefore, it is imperative that clinicians and researchers ensure the instrument is being administered as envisioned and validated, and that necessary precautions are being taken to minimize procedural infidelity.

The survey study reported here examines how the CAPE-V was being utilized across the United States primarily during the years 2009 through 2015. This exploratory study aimed to determine the extent to which practitioners strictly follow or how they might deviate from the published protocol for the CAPE-V. The results were expected to reflect how clinicians modify the tool in idiosyncratic ways to meet everyday time constraints and for other purposes. These findings may spur discussions regarding the impact of infidelity to the CAPE-V in clinical practice and research settings and have implications for possible changes to the 2009 published protocol that warrant consideration.

Auditory-perceptual assessment and the CAPE-V

Auditory-perceptual measures are the most common diagnostic tool used during a comprehensive voice evaluation and

Accepted for publication December 6, 2023.

From the *Department of Communication Disorders and Sciences, College of Health Sciences, Rush University, Rush University Medical Center, Chicago, Illinois; †Department of Neurology, Division of Speech-Language Sciences and Disorders, Henry Ford Health, West Bloomfield, Michigan; and the ‡Associate Professor, Emerita, Department of Communication Disorders and Sciences, College of Health Sciences, Rush University, Rush University Medical Center, Chicago, Illinois.

Address correspondence and reprint requests to: 888 S Michigan Ave Unit 900, Chicago, IL 60605. E-mail: gkempster@me.com

Journal of Voice, Vol xx, No xx, pp. xxx–xxx
0892-1997

https://doi.org/10.1016/j.jvoice.2023.12.009

are considered a gold standard for documenting voice disorders.[6] They are used to describe the sound quality of the voice, determine the presence and severity of a voice disorder, and aid in deciding the necessary course of treatment. While auditory-perceptual measurements lack the quantitative elements that acoustic and aerodynamic measurements offer, they remain a key standard for assessing vocal abnormalities.[1] In fact, a number of researchers argue that auditory-perceptual measurements provide the most salient information to both patients and clinicians. As it is typically the first sign of a vocal anomaly, an aberrant voice quality is often what compels the patient to seek professional treatment.[2]

During the last two decades, the CAPE-V gained popularity among clinicians and researchers in both clinical and laboratory settings.[3] However, its use in the laboratory setting can differ markedly from its use in the clinical setting.[7] The published CAPE-V protocol is often tailored in the research setting depending on the research question, the task at hand, and the availability of the patient or subject. Several research studies reveal partial use of the CAPE-V, such as utilizing exclusively the six CAPE-V sentences, one sustained vowel in lieu of the recommended two, or using the visual analog scale for measuring overall severity without rating the other five parameters.[8–11] For example, Vinson, Zraick, and Ragland conducted a perceptual analysis using a modified version of the CAPE-V in their experimental protocol; however, they did not provide additional details to explain what modifications were made and the rationale behind those modifications in their paper.[12] Use of the CAPE-V sentences exclusively for acoustic analysis has also been documented in research.[13] Lastly, the CAPE-V is being employed with research participants with primary diagnoses other than a voice disorder. Researchers have used the CAPE-V to assess the speech and voice in individuals with moderate-to-severe aphasia and apraxia of speech as well as in post-lingually deaf adults.[11,14]

While deviation from the CAPE-V protocol in the clinical setting has not been publicly documented, it is suspected that clinicians, similar to researchers, deviate from the specified protocol and in some way modify the assessment tool to meet everyday time constraints and provide individualized assessment. Occurrences like these are problematic given that the CAPE-V was designed for consistent and specified administration across all contexts.

The purpose of the investigation reported here was to determine the extent of procedural fidelity to the CAPE-V by surveying a small sample of practicing SLPs across the United States regarding their use of this tool. It is of critical importance to understand how clinicians are utilizing this assessment tool so that, if data suggest it is warranted, the instrument may be appropriately and thoughtfully revised to improve procedural fidelity and promote more consistent use in a variety of clinical and research settings.

METHODS

A comprehensive online survey was designed by the first author using an online survey tool (http://www.surveymonkey.com). The web-based survey incorporated a total of 59 questions. While the majority of questions were multiple-choice for analytical efficacy, two questions utilized a matrix grid, and six questions were open-ended to allow for free response. Many multiple-choice questions also included a comment box should participants want to elaborate on their responses. The questions on the matrix grid helped determine how often users of the CAPE-V administered each specific component of the evaluation tool. Each survey question targeted a different question domain, ranging from the general demographics of the participants to their preferences regarding the administration of the CAPE-V. Prior to releasing the survey, both authors answered all of the questions themselves to get an estimate of the time it would take for others to complete the survey. Appendix A includes a full list of the survey questions.

Potential participants for this pilot project were recruited via email through personal and professional contacts. All were licensed, practicing SLPs in the U.S. with current experience evaluating patients with voice disorders and were known to or familiar with the investigators involved in this study. The initial contacts were encouraged to forward the names and email addresses of any colleagues who utilize the CAPE-V in their practice and who might be willing to take approximately 15 minutes to complete this survey. A list of 19 potential participants who fit the inclusion criteria was compiled through this method.

The survey was disbursed electronically to these 19 individuals via an email that included a direct web link to the survey. All participants were given one week to complete the survey and were sent a reminder approximately 48 hours prior to the close of the study if they had not yet completed the survey. Responses included in this study were recorded between April 23, 2015 and April 29, 2015. Respondents were not compensated for participating in this online survey.

All respondents were given the option to leave questions blank or opt out of the survey at any time by closing the browser window on their electronic device. Following data acquisition, the data were compiled from the survey program and loaded into a spreadsheet program (Microsoft Excel).

RESULTS

Participants

Seventeen participants responded to the survey, yielding a response rate of 89.5% from the initial contact list. Basic demographic information is provided in Table 1. Participants represented 8 U.S. states, and included 12 women, 4 men, and 1 unspecified gender in answer to the question "Are you male or female?" All were current members of ASHA's Special Interest Group 3, Voice and Voice Disorders. Respondents' ages ranged from 20 seconds to 60 seconds, and years of clinical experience varied accordingly. Nine respondents identified themselves as clinicians, two as researchers, and six as both researchers and clinicians. All 17 participants reported that they evaluate and/or treat patients, clients, and/or subjects with suspected voice problems.

TABLE 1.
Respondent Demographics and Number of Years Using the CAPE-V

Demographic variable		Frequency (N = 17)	Percentage (100%)
Gender	Male	4	23.5
	Female	12	70.6
	No Response	1	5.9
Highest degree held	M.A.	4	23.5
	M.S.	9	52.9
	Ph.D.	3	17.6
	No Response	1	5.9
Professional role	Clinician	9	52.9
	Researcher	2	11.8
	Both	6	35.3
Years in practice	< 5	3	17.6
	6–10	1	5.9
	11–15	4	23.5
	16–20	3	17.6
	20+	5	29.4
Years using the CAPE-V	< 1	1	5.9
	1–3	1	5.9
	4–6	5	29.4
	6–10	7	41.2
	> 10	3	17.6

As expected, the majority of respondents reported having an academic background in voice and voice disorders. Five respondents reported taking one standalone course in voice disorders at the undergraduate level, and all of the respondents took at least one standalone course in voice disorders at the graduate level. Nearly all of the survey respondents reported taking continuing education (CE) courses in voice disorders and more than half of the respondents reported teaching, lecturing, and/or presenting CE courses related to voice.

Fifteen of the 17 respondents reported seeing voice-related patients on a weekly basis. The number of patients seen varied greatly: five respondents reported seeing 4–9 voice-related patients per week while another five respondents reported seeing more than 15 per week.

Respondents were asked to report if they used any methods or instruments to perceptually assess the voice other than the CAPE-V or the GRBAS scale. Only one respondent reported using another method, the Boone Voice Ratings Scales. All other respondents reported that they did not use any other perceptual analysis methods or left the question unanswered. The CAPE-V protocol and form[1] are included here as Appendix B and Appendix C for the reader to reference; copyright permission to include this information was granted from the American Speech-Language-Hearing Association.

Basic use of the CAPE-V

The length of time utilizing the CAPE-V varied by respondent (Table 1). The modal response was 6–10 years, selected by seven respondents. Another five respondents reported using the CAPE-V for 4–6 years. Additionally, respondents reported learning about the CAPE-V through a variety of methods as indicated in Figure 1. Respondents could choose more than one answer to this question.

All respondents reported having administered the CAPE-V in their practice (Figure 2). Sixteen (94.1%) reported currently using the CAPE-V for clinical purposes, research purposes, or both. Eight respondents (47.1%) reported having used or currently using the CAPE-V for research purposes, and 14 respondents (82.4%) reported having used or currently using the CAPE-V clinically. Fourteen respondents (82.4%) reported using the CAPE-V for 81–100% of their voice-related patients.

A majority of respondents reported administering the CAPE-V in less than 10 minutes, while two users reported taking 11–15 minutes to administer the tool. The frequency of administration varied as well: five respondents reported administering the CAPE-V multiple times a day, while six respondents reported administering the CAPE-V multiple times per week. One respondent skipped this question.

Given the number of components included in a comprehensive voice evaluation, SLPs must allot enough time to complete a thorough, informative assessment. Most respondents from the survey reported taking between 30 and 60 minutes to conduct a comprehensive voice evaluation, while two respondents reported taking more than 90 minutes to conduct a thorough voice evaluation (Table 2).

Procedural fidelity to the CAPE-V

The next series of survey questions inquired about the respondents' use of the CAPE-V in greater detail, examining the exact steps each individual takes when administering the tool.

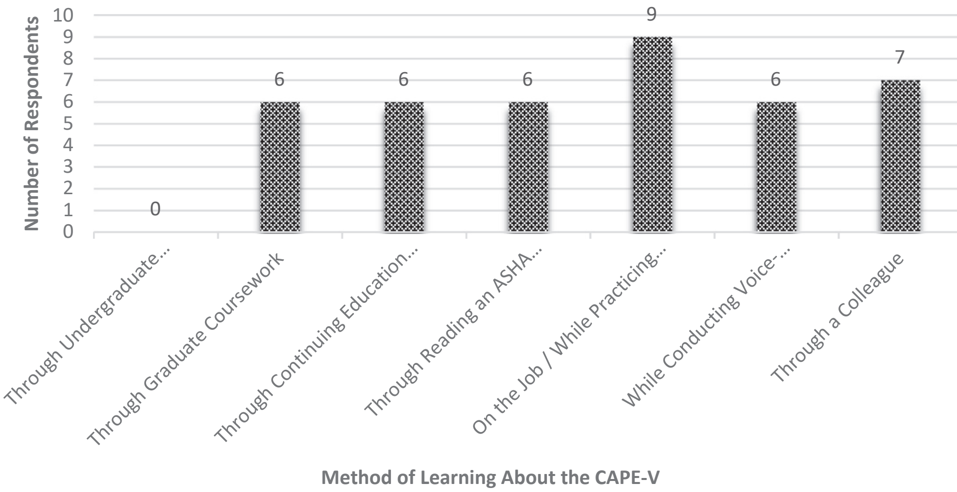

FIGURE 1. How respondents learned about the CAPE-V instrument.

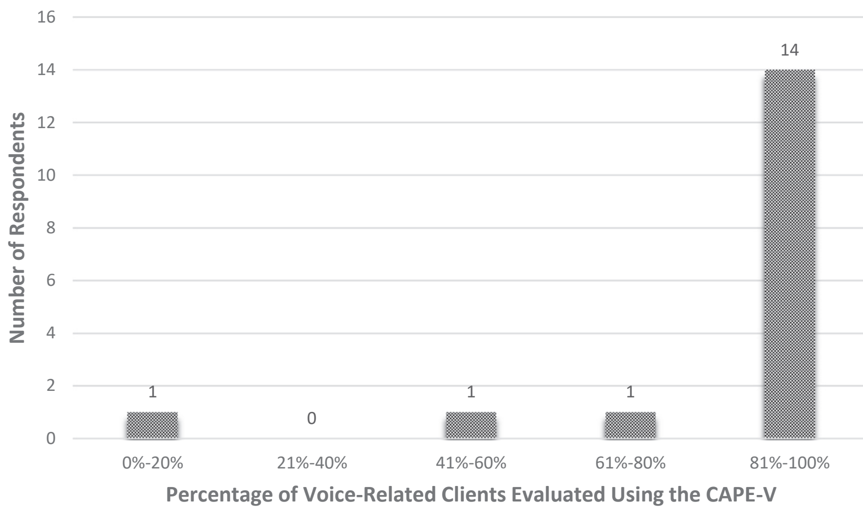

FIGURE 2. Percentage of voice-related clients evaluated using the CAPE-V.

Tasks

Fifteen respondents (88.2%) reported routinely using sustained vowels to administer the CAPE-V (Figure 3). Sixteen respondents (94.1%) reported routinely using sentences, and 16 respondents (94.1%) reported routinely assessing conversational speech. When examining those who deviated from the published procedure, one respondent reported only using conversational speech with the CAPE-V, while another reported skipping conversational speech altogether. A third respondent reported skipping sustained vowels. The survey further examined adherence to the smaller components of the CAPE-V protocol, such as the sustained vowel /i/ and each of the sentences. While a majority of respondents administered most of the specified CAPE-V tasks 81–100% of the time, only seven respondents (41.3%) reported administering all nine components of the CAPE-V. Notably, no single component of the CAPE-V was administered by all of the survey respondents 81–100% of the time.

Five respondents (29.4%) reported always following the exact task order as prescribed in the published protocol. Eleven respondents (64.7%) reported sometimes following the exact order. One respondent reported never following the written order or tasks, explaining that while she uses a minimum of conversational speech, she bases her ratings on whatever tasks she happens to hear while her students are working with patients.

Instructions

The set of instructions published along with the CAPE-V includes specifically worded instructions to use for each task when assessing the patient. Six survey respondents (35.3%) reported that they always use the instructions as given, and three respondents (17.6%) reported that they sometimes use them. Conversely, eight respondents (47.1%) reported that they never used the specifically worded instructions when administering the CAPE-V.

TABLE 2.
Respondents' Time and Frequency of Administration

Variable		Frequency (N = 17)	Percentage (100%)
Time taken to administer the CAPE-V			
# of Minutes	< 5	8	47.1
	5–10	7	41.2
	11–15	2	11.8
	16–20	0	
	> 20	0	
Time taken to complete a voice evaluation			
# of Minutes	< 30	1	5.9
	30–60	10	58.8
	61–90	4	23.5
	> 90	2	11.8
Frequency of CAPE-V administration			
Frequency	Multiple times per day	5	29.4
	One time per day	0	
	Multiple times per week	6	35.3
	Multiple times per month	2	11.8
	Multiple times per year	3	17.6

Vocal parameters

Although all respondents reported rating 'overall severity' at least 81–100% of the time while administering the CAPE-V, fewer respondents reported consistently rating 'roughness', 'breathiness', and 'strain' (Figure 4). Only 11 of the 17 respondents (64.7%) reported rating pitch and loudness the vast majority of the time (81–100%).

Sixteen respondents (94.1%) reported utilizing the option to add additional parameters to the rating form as needed (Table 3). While 16 reported that they have added additional prominent attributes to the blank scales on their CAPE-V rating form at some point in time, 12 respondents (70.6%) reported that they only add them 0–25% of the time. Some of the additional parameters commonly noted

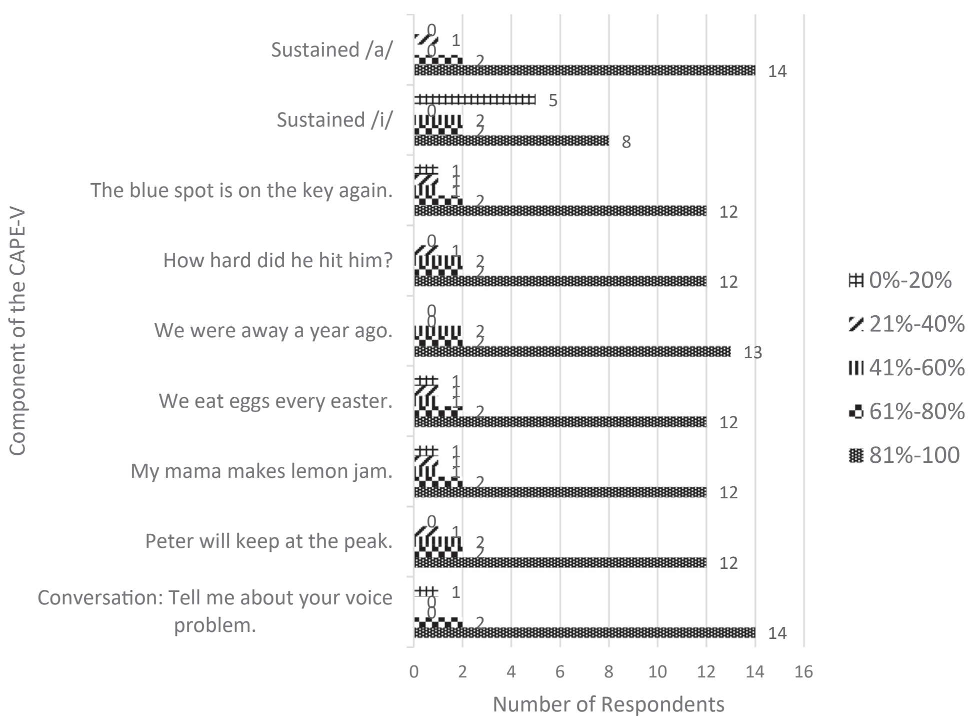

FIGURE 3. Percentage of time respondents administer components of the CAPE-V.

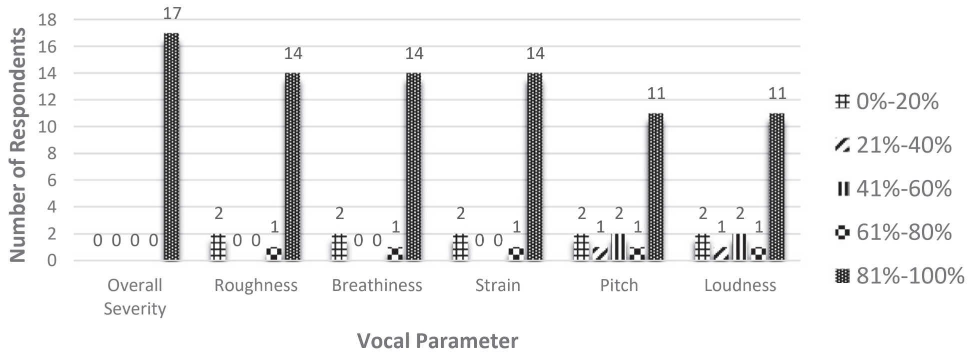

FIGURE 4. Average percentage of time respondents rate each vocal parameter on the CAPE-V.

TABLE 3.
Percentage of Time Respondents Add Additional Parameters

Variable		Frequency (N = 17)	Percentage (100%)
Percentage of time respondents add additional parameters			
% of time	0–25%	12	70.6
	26–50%	3	17.6
	51–75%	1	5.9
	76–100%	0	
	No response	1	5.9

by respondents included nasality, tremor or spasms, pitch or voice breaks, glottal fry, and diplophonia.

Consistency ratings

The indications of consistent or intermittent presence for each vocal attribute on the CAPE-V rating form were reported. Thirteen respondents reported they always, three respondents reported they sometimes, and one respondent stated that she does not ever classify the vocal parameters as consistent or intermittent. Some respondents elaborated on this task and revealed that they only flag vocal attributes when they are intermittent or inconsistent, not when they are consistently present.

In order to use the CAPE-V rating scale, users are instructed to place a tick mark on each of the visual analog scales while listening to each task type. Sixteen respondents (94.1%) reported that they always place tick marks on the VAS when administering the CAPE-V, while one respondent (5.9%) reported sometimes placing tick marks on the VAS. In terms of measurement, 13 respondents (76.5%) reported always using a ruler to measure the exact placement on the scale; three respondents (17.6%) reported that they try to use a ruler but it is not always available, so they estimate, or they neither estimate nor use a ruler. Follow-up responses revealed that four respondents (23.5%) utilize an electronic version of the CAPE-V, which assists them in measuring the exact placement of the tick mark.

Finally, all respondents were asked about the tick mark protocol listed in the CAPE-V procedures and instructions. In general, the tick mark protocol specified for the CAPE-V is somewhat complicated and requires that users apply multiple tick marks and label each tick mark with the corresponding task number if they hear discrepancies across or between task types.

Only three respondents (17.6%) reported that they always follow the tick mark protocol as written, while nine respondents (52.9%) reported that they sometimes follow it as written. The remaining five respondents (29.4%) reported that they do not follow the tick mark protocol as written when administering the CAPEV. One respondent explained that she has not used the instrument enough to practice using the tick mark protocol, and other respondents reported that they do not follow this specified tick mark protocol unless there are major, perceptually variant differences between task types. One respondent indicated that a computerized program for the CAPE-V does not allow users to follow the tick mark protocol as written.

The CAPE-V protocol instructs users to audio record their patients' performances on each of the three task types using standard recording procedures.[1] Survey responses revealed that 13 respondents (76.5%) always audio record their patients, three respondents (17.6%) sometimes audio record their patients, and one respondent (5.9%) does not audio record patients.

Survey comments: issues and variance in practice

The final portion of the CAPE-V survey included a series of open-ended questions. These questions aimed to better understand the preferences and opinions of CAPE-V users as well as determine any preferential patterns that may have developed since the inception of this measurement tool. These free-response questions were optional, therefore not all respondents provided feedback for this section of the survey. Table 4 for a compilation of the responses to the four open-ended questions, presented below.

TABLE 4.
Responses to Open-ended Survey Questions

Open-ended questions		Responses
Q54	Are there aspects of the CAPE-V protocol and rating form that you do not routinely incorporate into your practice? If so, what, and why?	"I do not audio record my CAPE-V samples, primarily due to convenience. It is cumbersome to add these recordings and impacts the efficiency of evaluations." "Strain rating is included, but is always a source of disagreement between clinicians." "I do not specifically have the person follow the tasks and do the ratings as separate procedures. The ratings occur throughout the evaluation, as the patient is engaging in speech tasks for other purposes." "I don't always do the sentences." "I do not regularly incorporate all the sentences. I only use them for differential diagnosis of [spasmodic dysphonia]. If vocal quality is the same on reading and connected speech, I don't feel it is necessary to take the time for the sentences." "I don't routinely 'earball' pitch and loudness, because they are quantified as part of my assessment. I'll measure or note them perceptually with the CAPE-V if they are very abnormal, but for many clients I ignore those two lines."
Q55	Are there aspects of the CAPE-V protocol and form that you find especially helpful? If so, what, and why?	"Different tasks to highlight different settings in which dysphonia may be more or less apparent." "I appreciate the standardization of the voice quality dimensions and the use of the visual analog scales." "Ability to rate different perceptual features of voice, their consistency across a variety of tasks, and then use that scale for pre- and post-testing." "Use of different kinds of sentences (e.g. all voiced vs. voiceless) to help differentiate severity in some contexts (e.g. spasmodic dysphonia)." "Contrasting the sustained vowel with connected speech and reading. Often the quality on the sustained is different from on or both of the speaking tasks." "As mentioned before, the inclusion of pitch in the CAPE-V is helpful."
Q56	In your opinion, what aspects of the CAPE-V protocol and/or form should be changed to make it easier and more efficient to use?	"Reconsider the need to audio record all samples." "Computerized version would be great." "Some clinicians feel that there are rules for how the various dimensions should "add up" (e.g. overall severity should be equal to the highest rated sub-dimension). To my understanding, this is not a criterion of the tool (as roughness could be rated higher than overall severity). It would be nice to clarify this." "Automate (and make the app free) so that a ruler isn't necessary." "There are two versions. It should be clear that using the one with equally placed text anchors may provide different results than using the one with unequally placed anchors." "The individual rating of each element is not necessary unless there are significant changes in one or more. I do not use all the sentences at one time, in part because they do not contribute additional information to my assessment of the voice unless I suspect [spasmodic dysphonia] or I feel I need to isolate certain behaviors."

TABLE 4. (*Continued*)

Open-ended questions		Responses
Q57	Do you have any other suggestions for modifications of the CAPE-V?	"Make the electronic version more easily available, particularly on tablets or smartphones. Create PDFs of the resulting evaluation form for download to the EMR." "Having scoring software would be nice." "It should be available electronically." "Provide anchors and training materials." "I would change 'How hard did he hit him' due to content and 'Peter will keep at the peak' because patients struggle with the atypical wording." "See CAPE-V validation study by Zraick et al. for suggested changes and modifications."

The first question targeted aspects of the CAPE-V that users do not routinely incorporate into their practice. Two respondents mentioned that they do not incorporate all or any of the sentences, particularly if vocal quality sounds the same during connected speech and during a reading task. One respondent reported that she does not audio record the CAPE-V samples, as it is "cumbersome" and "impacts the efficiency of evaluations." Another respondent revealed that she ignores the visual analog scales for pitch and loudness, as they are quantified in other parts of her comprehensive voice assessment, such as the acoustic component.

The second free-response question allowed respondents to disclose aspects of the CAPE-V that they find especially helpful. In contrast to a response to the previous question, one respondent revealed appreciation for the inclusion of pitch in this tool. Several respondents appreciated the variety of tasks and sentence types that highlight vocal quality in multiple phonetic contexts. Another user commented on the use of the visual analog scales and the standardization of the voice quality dimensions.

The third open-ended question inquired about aspects of the CAPE-V protocol and form that should be changed to make it easier and more efficient to use. Most responses to this question alluded to automating the CAPE-V to an appbased or computer-based software because an electronic version that generates reports could streamline scoring and documentation. One respondent commented on reconsidering the need to audio record the CAPE-V tasks, and another respondent commented on clarifying for users whether the visual analog scales need to add up in the overall severity rating or if overall severity can be rated less than the other scales included on the rating form. Another respondent brought up the fact that there are two versions[1] of the CAPE-V form in circulation, one with equally placed, centered text descriptors (mild, moderate, severe) below the VAS and one with skewed text descriptors. This respondent stressed the need to communicate the discrepancy between the two versions to the voice community, given that using one or the other may provide different results for CAPE-V users.

A fourth open-ended question asked respondents to provide suggestions for modifications of the CAPE-V. One respondent reported that anchors and training materials may be helpful for those learning to use the tool. Another respondent would change Sentence 2 (ie, "How hard did he hit him?") and Sentence 6 (ie, "Peter will keep at the peak") due to content and atypical wording. The last respondent referred to the CAPE-V validation study published by Zraick et al (2011).[2] Some suggested recommendations from that study include specific training of CAPE-V procedures and use of the scales as well as possibly using anchor stimuli.

DISCUSSION

The intent of this pilot survey study was to better understand how clinicians and researchers utilize the CAPE-V and whether they follow the published protocol, thereby reflecting procedural fidelity to the instrument as intended. It is evident from previous literature that researchers who use the CAPE-V have been deviating from the standardized, published protocol outlined in Kempster et al, 2009.[1] Results of the survey reported here not only confirmed many of these deviations in clinical as well as research settings but add insight into how and why users are deviating from the published CAPE-V protocol. It is important to emphasize that when users are not faithful to established procedures, reliability is at risk, and validity is threatened for any instrument.[15]

Deviations from fidelity

The first deviation from procedural fidelity pertained to consistency in CAPE-V protocol administration. Half of the respondents reported that they were either unsure if their personal protocol for this instrument matched that of their colleagues or that they were aware that they did not administer the CAPE-V in the same way as their colleagues. This not only suggests that users are deviating from

[1] The original ("beta") version of the CAPE-V form released by SIG 3 in 2002 had positively skewed visual descriptors (ie, mild, moderate, severe) under each VAS. However, the published 2009 version of the CAPE-V form has equally spaced, centered visual descriptors under each VAS.

the original, published protocol of the CAPE-V, but also shows that users within the same clinical practice are likely using the CAPE-V differently from one another. This finding has a potentially serious effect on inter-rater reliability and could skew pre-and post-test results of a patient or subject.

A second deviation was that respondents did not always administer all three task types when using the CAPE-V. ***In fact, no single user reported exactly maintaining fidelity to the CAPE-V as specified in the published protocol.*** For example, only half of the participants who responded consistently asked patients to sustain the vowel /i/ when administering the CAPE-V, and several respondents reported administering some, but not all, of the CAPE-V sentences. This latter point may reflect additions to personal CAPE-V protocols, such as the inclusion of a reading (ie, the Rainbow Passage). One user reported not using all of the sentences if there are no discrepancies during an oral reading task (a task not a part of the CAPE-V).

Lastly, only five respondents reported always following the published task order, beginning with sustained vowels and ending with conversational speech. Some users reported adding the Rainbow Passage, while other respondents reported beginning with conversational speech, as it is logical to have a conversation with patients or subjects when discussing their voice problems for the first time.

The most surprising deviation was found through survey questions which revealed that only 6 out of the 17 respondents claimed they always use the specifically worded directions written in the original, published protocol. Deviating from the specifically worded instructions increases the chance for variation across patients, clinicians, and centers and reduces procedural fidelity. Moreover, only 3 out of the 17 respondents reported always following the tick mark protocol as indicated in the published instructions, which may result in failing to document aberrations in voice quality in certain task types.

Perhaps some of these deviations from procedural fidelity are a result of a lack of education regarding the proper use of the CAPE-V tool. Further training may be needed for those SLPs and other professionals who use the CAPE-V. Time restrictions may also play a role in these deviations, given that a majority of respondents indicate they administer the CAPE-V in less than 5 minutes. In reality, one would expect it might take more than 5 minutes to administer all of the specific components of the tool when following the published protocol, especially if an examiner chooses to document the findings throughout the assessment, as opposed to listening to recorded samples and scoring a voice later.

Given that the participants of this survey study were recruited through personal and professional contact, they may represent avid CAPE-V 'super-users' and not the general population of SLPs. Thus, a few trends were noted in this survey that may not represent those of a typical voice clinician. Nearly half of the respondents identified themselves as researchers or as both researchers and a clinician and reported using the CAPE-V for research purposes. However, with a sizeable proportion of respondents using the CAPE-V for research purposes, one might expect greater adherence to the published protocol for this tool, and our data do not seem to support this notion. It is likely that a broader survey with a greater number of respondents would result in a substantial change in the proportion of clinicians versus researchers who use the CAPE-V; questions exploring the use of the CAPE-V in each of the different roles may provide data of considerable interest.

Overarching themes and patterns

Given the trends noted here, it is likely that a larger, updated survey would show even less adherence to the CAPE-V protocol (ie, more variation in usage) than indicated by this study. One future direction for the CAPE-V appeared clear—the need for an electronic version. While an electronic version of the CAPE-V was not widely available at the time of this study, the overarching theme resulting from the qualitative, open-ended responses of this survey study is for automation—that is, for implementation of an electronic version of the CAPE-V, which can be easily incorporated into other computerized aspects of a full evaluation. The future of the CAPE-V would ideally be a computer-based, automated program, or one that is perhaps app-based for use on smartphones and tablets. Importantly, an automated version of this auditory-perceptual assessment tool may help to increase adherence to the original, standardized protocol without needing to change the fundamental CAPE-V principles. Some obstacles to this change in methodology are obvious: First, the application must be approved for use by ASHA, which holds the copyright for the CAPE-V. Second, the program or app would have to be able to automatically audio record each of the CAPE-V tasks for later use and store them in a device or on the cloud for documentation and review. Third, the program would ideally include a pop-up window with the specifically worded instructions, and in a specific order, so users can conveniently read off of the device to administer each task. Finally, potentially the app can mandate that users indicate the location of a tick mark for each task type, compelling clinicians to adhere to the protocol as written. Of course, issues regarding the security and confidentiality of patient information and transfer of information to a patient's medical record would have to be resolved.[2]

Because respondents were given the option to leave questions in the present survey unanswered, the open-ended questions received the least amount of attention and had the

[2] In 2018, Pentax Medical released an electronic version of the CAPE-V, called the "iCAPE-V" bundled with their Visipitch instrument, a widely used clinical evaluation tool. Users report that the iCAPE-V does address some of the issues raised here.

fewest number of overall responses. Besides the several comments regarding an electronic version of the CAPE-V, no other consistent themes were found when examining the open-ended responses. However, it is likely that more patterns would arise with a larger group of respondents.

Limitations

This study had significant limitations, some of which warrant further discussion. There is no doubt that this exploratory study unearthed a number of deviations from procedural fidelity to the CAPE-V. However, a secondary purpose of this investigation at the time it was conducted was to potentially develop a survey instrument suitable for use on a national basis. One example of a small, but updated change to a question on the survey, would be to change the options for the response regarding gender identity from the simple binary “male or female” choice used here to include other, more inclusive and diverse options. With respect to the survey data as a whole, the sole use of frequency counts and percentages to analyze the data limits our understanding of the strength, potential, and effects of the results reported here. A broader survey, based in the science of survey development, could/should be drafted in such a way as to demonstrate internal consistency of items and so that responses within and between each domain could be explored. Furthermore, the data obtained through a revised survey could be set up to allow for correlational measures and more robust statistical analyses. This would provide insight into the covariation present across the variety of questions and responses and perhaps point to clear associations among the variables.

The analyses using in this study were purely descriptive, restricted by the survey design, the sample size, and the diversity of responses obtained. Not enough respondents were targeted to form groups for comparison, with demographic data and behavioral data showing large ranges and high variation. However, the data that were obtained here would be useful to refine the survey to develop a more robust instrument. This would result in improved validity of this method to evaluate the current use of the CAPE-V and to assess evident patterns or trends.

FUTURE RESEARCH AND DEVELOPMENT

As EBP patterns become the norm in the field of speech language pathology, the construct of fidelity will become even more clinically relevant. While achieved fidelity in practice will never match intended fidelity in the laboratory setting, maintaining fidelity is critical to minimize any external variables that may decrease the overall reliability and validity of an assessment or treatment protocol. This is especially true for an instrument such as the CAPE-V where the goal is to document perceived auditory-perceptual voice quality. The study reported here revealed a number of variables that affect the procedural fidelity to the CAPE-V by experienced users. Therefore, our results suggest that a revision of the CAPE-V appears warranted, based on available evidence of user strategies and considerable variability in practice. A revision of the CAPE-V would likely improve the ease and consistency with which both clinicians and researchers use this assessment tool. Any such revision could be undertaken with attention paid to ways to improve procedural fidelity as well as to take into consideration current practices in health care with respect to efficiency, productivity expectations, and documentation requirements.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gail B Kempster reports financial support was provided by Rush University. Gail B Kempster reports a relationship with Rush University that included: employment. The corresponding author has previously been published in the Journal of Voice and has served as a reviewer. These activities occurred more than 5 years ago. GBK.

Acknowledgments

The first author would like to thank Mrs. Mary Bacon, Dr. Richard Peach, and Mrs. Laura Vygantas for serving as members of her thesis committee and for providing constructive feedback and advice during this project. Both authors are also grateful to Dr. Kathleen Nagle (Seton Hall University) and Dr. Nancy Solomon (Walter Reed National Military Medical Center) for their editorial suggestions. Special thanks are extended to Isabella Toral who assisted the authors with manuscript formatting. Finally, the authors extend their appreciation to all of the survey participants for their time, energy, expertise, and recommendations for the CAPE-V. The data reported here were presented as a poster at the American Speech-Language-Hearing Association Convention in Denver, Colorado (November 2015).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jvoice.2023.12.009.

References

1. Kempster GB, Gerratt BR, Abbott KV, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol.* 2009;18:124–132. https://doi.org/10.1044/1058-0360(2008/08-0017).
2. 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. https://doi.org/10.1044/1058-0360(2010/09-0105).

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. Hulleman CS, Cordray DS. Methodological studies: moving from the lab to the field: the role of fidelity and achieved relative intervention strength. *J Res Educ Eff.* 2009;2:88–110. https://doi.org/10.1080/19345740802539325.
5. Kaderavek JN, Justice LM. Fidelity: an essential component of evidence-based practice in speech-language pathology. *Am J Speech Lang Pathol.* 2010;19:369–379. https://doi.org/10.1044/1058-0360(2010/09-0097).
6. Oates J. Auditory-perceptual evaluation of disordered voice quality. *Folia Phoniatr Logopaed.* 2009;61:49–56. https://doi.org/10.1159/000200768.
7. 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.
8. Gillespie AI, Gooding W, Rosen C, et al. Correlation of VHI-10 to voice laboratory measurements across five common voice disorders. *J Voice.* 2014;28:440–448. https://doi.org/10.1016/j.jvoice.2013.10.023.
9. 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).
10. Dietrich M, Abbott KV. Vocal function in introverts and extraverts during a psychological stress reactivity protocol. *J Speech Lang Hear Res.* 2012;55:973–987. https://doi.org/10.1044/1092-4388(2011/10-0344).
11. Ubrig MT, Goffi-Gomez MVS, Weber R, et al. Voice analysis of postlingually deaf adults pre- and postcochlear implantation. *J Voice.* 2011;25:692–699. https://doi.org/10.1016/j.jvoice.2010.07.001.
12. Vinson KN, Zraick RI, Ragland FJ. Injection versus medialization laryngoplasty for the treatment of unilateral vocal fold paralysis: follow-up at six months. *Laryngoscope.* 2010;120:1802–1807. https://doi.org/10.1002/lary.20982.
13. Stepp CE, Merchant GR, Heaton JT, et al. Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyper function. *J Speech Lang Hear Res.* 2011;54:1260–1266. https://doi.org/10.1044/1092-4388(2011/10-0274).
14. Kurland J, Pulvermüller F, Silva N, et al. Constrained versus unconstrained intensive language therapy in two individuals with chronic, moderate-to-severe aphasia and apraxia of speech: behavioral and fMRI outcomes. *Am J Speech Lang Pathol.* 2012;21:S65–S87. https://doi.org/10.1044/1058-0360(2012/11-0113).
15. Nagle K. Clinical use of the CAPE-V scales: agreement, reliability, and notes on voice quality. Journal of Voice. doi:10.1016/jvoice.2022.11.014.