

French-Language Adaptation of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)

*†Timothy Pommée, ‡Déborah Mbagira, and ‡Dominique Morsomme, *†Toronto, Canada, and ‡Liège, Belgium

Summary: Objectives. This study aimed to adapt the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) protocol for perceptual voice assessment to the French language. The primary objective was to achieve consensus among an international panel of voice experts on the content of the adapted protocol.

Methods. To ensure the relevance and robustness of the French CAPE-V protocol, this study employed a systematic Delphi method and involved an international panel primarily comprising speech therapists and lecturers from France and Belgium. The multi-stage process included an initial panel size of 15 experts. Three rounds of online questionnaires, integrating both quantitative and qualitative data collection, were conducted. Participants provided feedback and ratings on various protocol elements until a consensus was reached. Adaptations targeted the choice of task stimuli (sustained vowel, sentence reading, semi-spontaneous speech), of the rating scales, and vocal quality terminology.

Results. The Delphi process achieved consensus on all elements of the adapted CAPE-V protocol. Notably, the sustained vowel task saw consensus in favor of the vowel /a/. Sentence adaptations achieved substantial agreement, with the final set unanimously approved. The simple Visual Analog Scale emerged as the preferred rating scale. Agreement on terms for describing vocal qualities marked a crucial step in establishing a shared vocabulary among French-speaking voice experts.

Conclusions. The study successfully adapted the CAPE-V protocol for perceptual voice assessment to the French language through a systematic Delphi process. The final protocol closely resembles the original English version, maintaining its structure and core objectives. Consensus on sustained vowel tasks, sentence adaptations, rating scales, and vocal quality terminology indicates the relevance and robustness of the adapted protocol. Ongoing validation studies in France demonstrate the potential clinical utility of the adapted CAPE-V in French-speaking contexts, representing a significant step toward standardized and validated voice assessment tools for clinicians and researchers globally.

Key Words: Voice–Assessment–Evaluation–Auditory-perceptual–CAPE-V–French.

INTRODUCTION

The standard clinical assessment of voice disorders involves a comprehensive patient history, acoustic and aerodynamic measures, as well as perceptual and self-assessment tools.¹ Their combination allows to quantify and qualify the patient's vocal behavior in various social, private, and professional life situations to suggest an appropriate therapeutic plan.²

The use of auditory-perceptual tests remains the gold standard of the clinical voice assessment. It is based on speech samples produced following specific instructions on various vocal materials (eg, sustained vowel, spontaneous speech, singing voice, reading passage). The reliability of perceptual measurements depends on the scoring method, the stimulus type, as well as clinician-related factors.³ Lee et al⁴ demonstrated that auditory-perceptual evaluations have high inter- and intra-rater reliability when anchor

voices are presented (eg, a voice sample considered as normophonic) before assessing pathological samples. Despite the well-known limits relating to their reliability, perceptual ratings are non-invasive and remain the most clinically relevant and ecologically valid tool for analyzing the voice, which is an inherently perceptual phenomenon.⁵

The most common perceptual assessment tool in Europe is the GRBAS (Grade, Roughness, Breathiness, Asthenia, and Strain) scale.⁶ Created by Hirano in 1981,⁷ it describes vocal quality based on five vocal parameters, rated from zero (normal) to three (severe): Grade (overall voice quality), Roughness, Breathiness, Asthenia, and Strain. This multi-dimensional scale is widely used in clinical practice and research to assess the severity of all types of dysphonia,⁸ either on conversational speech or sustained vowels.⁹ It has the advantage of being easy to use and quick to administer, and its reliability and validity have been studied.⁸ However, the GRBAS scale also has important limitations. To begin with, not all evaluation criteria are equal. While the overall assessment of vocal quality (G) achieves the best intra- and inter-rater agreement,¹⁰ the perception of strain and astheny shows low reliability.¹¹ Furthermore, Ghio et al⁶ reported low intra-rater reliability and questioned the sensitivity of the G, B, and R criteria, as highly dysphonic voices were observed to be easier to evaluate than voices described as “intermediate” or borderline pathological. Another limitation of this tool is the absence of a standardized protocol,

Accepted for publication March 12, 2024.

From the *Sunnybrook Research Institute, Toronto, Ontario, Canada; †Department of Speech-Language Pathology, Rehabilitation Sciences Institute, University of Toronto, Toronto, Ontario, Canada; and the ‡Research Unit for a Life-Course Perspective on Health and Education, Université de Liège, Liège, Belgium.

Address correspondence and reprint requests to Timothy Pommée: 500 University Avenue, Toronto, ON M5G 1V7, Canada. E-mail: timothy_pommee@hotmail.com
Journal of Voice, Vol xx, No xx, pp. xxx–xxx
0892-1997

© 2024 The Authors. Published by Elsevier Inc. on behalf of The Voice Foundation. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). <https://doi.org/10.1016/j.jvoice.2024.03.011>

which would greatly improve both intra- and inter-rater reliability. Finally, this scale demonstrates a low sensitivity due to the nature of its scoring.⁸

Another commonly used tool to assess voice disorders is the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V).¹² The CAPE-V was developed during a conference organized by the American Speech-Language-Hearing Association (ASHA), following the observation that clinicians lacked standardized, reliable, and scientifically validated perceptual analysis tools, making communication among clinicians challenging. In addition to describing the severity of auditory-perceptual characteristics, the CAPE-V helps formulating anatomical and physiological hypotheses related to the voice disorder and assists in the decision of whether additional tests are needed. The CAPE-V assesses six salient perceptual characteristics of the voice: overall severity, roughness, breathiness, tension, pitch, and loudness. The clinician rates the degree of each characteristic on a Visual Analog Scale (VAS, a 100-mm line), aided by descriptors on the scale (eg, “mild” indicating a “slightly deviant” characteristic). The clinician can also specify whether a particular vocal characteristic is consistently present or intermittent. Recognizing the non-exhaustive nature of the criteria, the authors offer two blank VAS that the clinician can label and rate based on their own observations.

In the original English-language protocol, the evaluation begins with a sustained vowel task (/a/ and /i/) to observe vocal quality in isolation, without any effects relating to continuous speech.⁹ The second task involves reading aloud six sentences designed to elicit specific oro-laryngeal behaviors. For instance, the first sentence is said to contain all English vowel sounds, the fifth incorporates nasal sounds, and the last sentence contains voiceless plosives. During the third and final task, the clinician assesses the patient's use of their voice in semi-spontaneous speech, using the standard questions “Tell me about your voice problem” or “Tell me how your voice is functioning.”

While the CAPE-V suffers from the same main limitation as all examiner-based perceptual tools with regards to its subjectivity, it has the advantage of having a precise standardized administration protocol, increasing its reliability. Furthermore, the VAS makes the tool more sensitive and the scoring more precise—yielding data that are more convenient for statistical analysis¹³—and achieves better reliability than a Likert scale.^{6,8} The use of various tasks (sustained vowel, sentence reading, and spontaneous speech) allows for a comprehensive evaluation of laryngeal motor behaviors. Recently, research has extensively tested this tool, comparing it, notably, to the GRBAS as a reference perceptual scale for evaluating dysphonic voices.^{8,11,14,15}

The CAPE-V has been adapted and validated in at least 12 languages.^{16–29} These adaptation efforts have been achieved through different methods and expert collaborations, as no standard procedure exists to this end. This methodological heterogeneity has recently been confirmed in a systematic review focusing on the methods used for adaptation and

validation of the CAPE-V protocol.³⁰ For example, for the Hindi version,¹⁸ a linguist expert matched the number of target sounds in the original sentences to create new sentences. In Spanish,¹⁷ the protocol was phonetically adapted by a speech therapist. In Malay,¹⁶ the protocol was adapted by a team consisting of three bilingual English-Malay speech therapists and one linguist. They implemented a forward-backward translation system, where the initial translation was done by a speech therapist and a linguist. Subsequently, the linguist analyzed the translation for phonemic consistency with the original CAPE-V. In the second phase, the other two speech therapists retranslated the tasks without referring to the original protocol. Finally, all translators convened to discuss items they had translated differently. The form itself, such as the use of the VAS and its three descriptors, is generally retained in most adaptations. Only the Mandarin adaptation²² slightly increased the length of the VAS to facilitate scoring. A detailed review of the various methods used to adapt the CAPE-V into several languages is given elsewhere.³⁰

To date, despite its psychometric advantages over the GRBAS scale, the CAPE-V assessment tool is not available to French clinicians. To adapt the CAPE-V tool to the French language and to maximize its future implementation in clinical practice, we chose to make use of the Delphi consensus method. The aim of this study is to describe the adaptation process of the CAPE-V protocol into the French language, to provide a protocol tailored to the French language as it is spoken in Belgium, France, and other French-speaking regions, accessible for both clinical practice and research. To that aim, a consensus needs to be achieved regarding three main points for the French adaptation of the CAPE-V: 1) the stimuli used to elicit voice samples in the sustained phonation, sentence reading, and semi-spontaneous speech tasks, 2) the most suitable rating scale for perceptual voice assessment on these tasks, and 3) the terminology used to describe vocal characteristics such as “roughness,” “breathiness,” and “strain” from the original protocol. More than just proposing the French protocol, we wish to explicit the method used to achieve a thorough consideration of both the form and content of each part of the protocol, involving an international expert panel. We hope this will be of use to other authors in the future wishing to adapt the protocol to their respective languages.

METHODS

Ethics approval

This research received approval from the Ethics Committee within the Faculty of Psychology, Speech Therapy, and Education at the University of Liège (agreement n°: 2022-5453). All individuals whose data contributed to this study provided informed consent for their data to be utilized for research purposes. Participants granted permission for their responses to be employed in an anonymized and grouped fashion for the creation of consensus statements. The only personal information known to the survey moderator was the participants' email addresses, which facilitated the monitoring of the survey across multiple rounds.

Study design

The Delphi method is a structured, iterative, and interactive multi-stage process that allows for the involvement of large international expert panels.^{31,32} It is generally conducted over several consecutive rounds (usually three^{33–35}). In a “modified Delphi study,” initial statements are presented in the first round, derived from a preliminary literature review.^{36,37} Following each round, the panel’s responses are synthesized, and areas of consensus or disagreement are identified. Aggregated controlled feedback³⁸ is then provided to the panel in the subsequent round to clarify alterations made to facilitate consensus, and participants provide their input on the new statements.^{37,39} These iterations continue until consensus is achieved or until stability in responses is attained (see^{38,40} for examples of stability measures).

The Delphi process was chosen because it aligns with our goal of obtaining a relevant adaptation of the CAPE-V in French. Among its numerous advantages, the Delphi method offers cost-effectiveness and surpasses geographical constraints, as rounds can be conducted via mail or online. It thereby allows to include a large number of participants, enhancing the validity of the obtained results.³⁵ For the present study, the participation of voice experts from Belgium, Quebec, the Maghreb region, Switzerland, and France is particularly valuable in achieving an inclusive adaptation. Another significant benefit is its quasi-anonymous nature.^{38,41} Participants’ identities remain undisclosed to one another, fostering open expression free from social or professional peer pressure. The quasi-anonymous nature, combined with the use of multiple rounds and structured feedback, helps mitigate bias in the pursuit of consensus.⁴⁰ Given these advantages, the Delphi method was selected for this study to collaboratively adapt the CAPE-V protocol into French with experts who will also be the eventual users of this assessment tool.

The stages of the present modified Delphi study are summarized in [Figure 1](#). The main steps that will be described hereafter are the expert selection, initial draft of the French protocol, and construction of the three consecutive Delphi rounds.

Definition of the problem

The objective of the present study is to achieve a consensus regarding three main points for the adaptation of the CAPE-V into French: 1) the stimuli used to elicit voice samples in the sustained phonation, sentence reading, and semi-spontaneous speech tasks, 2) the most suitable rating scale for perceptual voice assessment on these tasks, and 3) the terminology used to describe vocal characteristics, such as “roughness,” “breathiness,” and “strain” from the original protocol. Our primary research hypothesis is that the resulting consensual adaptation of the CAPE-V protocol will be relevant for use in various French-speaking

countries (Canada, Belgium, France, the Maghreb region, and Switzerland).

Selection and recruitment of experts

We targeted four professions of voice experts: vocologists, phoneticians, speech therapists, and otorhinolaryngologist. Participants were required to have a minimum of 5 years of clinical or research experience in their field.

As the adaptation of the protocol required participants to have an excellent knowledge of the French language, all participants were required to be native French speakers. Clinicians were required to work with a French-speaking patient population to be aware of the practical challenges of adaptation.

All participants who did not meet the aforementioned criteria were excluded from this study.

We aimed to recruit a panel of 30 individuals, based on previous recommendations to limit the impact of non-responses from participants on the reliability of the results.³²

Recruitment was carried out via national professional associations and private groups on social media. Additionally, individual emails were sent to hand-selected voice experts identified in literature searches on PubMed, who had at least three publications in the field of voice, were authors of a reference book or participated in research projects linked with voice pathology. Overall, more than 200 individuals and institutions were contacted by email and notices were posted on 19 groups linked to speech therapy in French.

Non-respondents for each round were excluded from subsequent rounds. An a posteriori analysis using descriptive statistics was carried out to verify that the expert panel characteristics did not change.

A total of 16 experts completed round 1; 9 experts completed round 2; and 8 experts completed round 3.

Initial draft of the French protocol

The three tasks from the original protocol (sustained vowel, sentences, and induced speech) and their original aims were retained and adapted to allow the observation of relevant laryngeal behaviors on samples in French. The present study used a modified Delphi process, as an initial version of the adaptation of the protocol was drafted based on a literature review (in-depth analysis of the original CAPE-V materials and review of study methods for translation into other languages) and a collaboration with a phonetician for the adaptation of the six sentences. The phonetician provided input on phonetic and syntactic coherence with regards to each sentence-specific aim. This initial draft was then subjected to the Delphi process.

At the time of launching the first round of our Delphi survey, we discovered that an unpublished adaptation of the CAPE-V into Quebec French was made by the Community of Practice in Voice Therapy of Quebec. Indeed, a team of speech therapists from Quebec convened

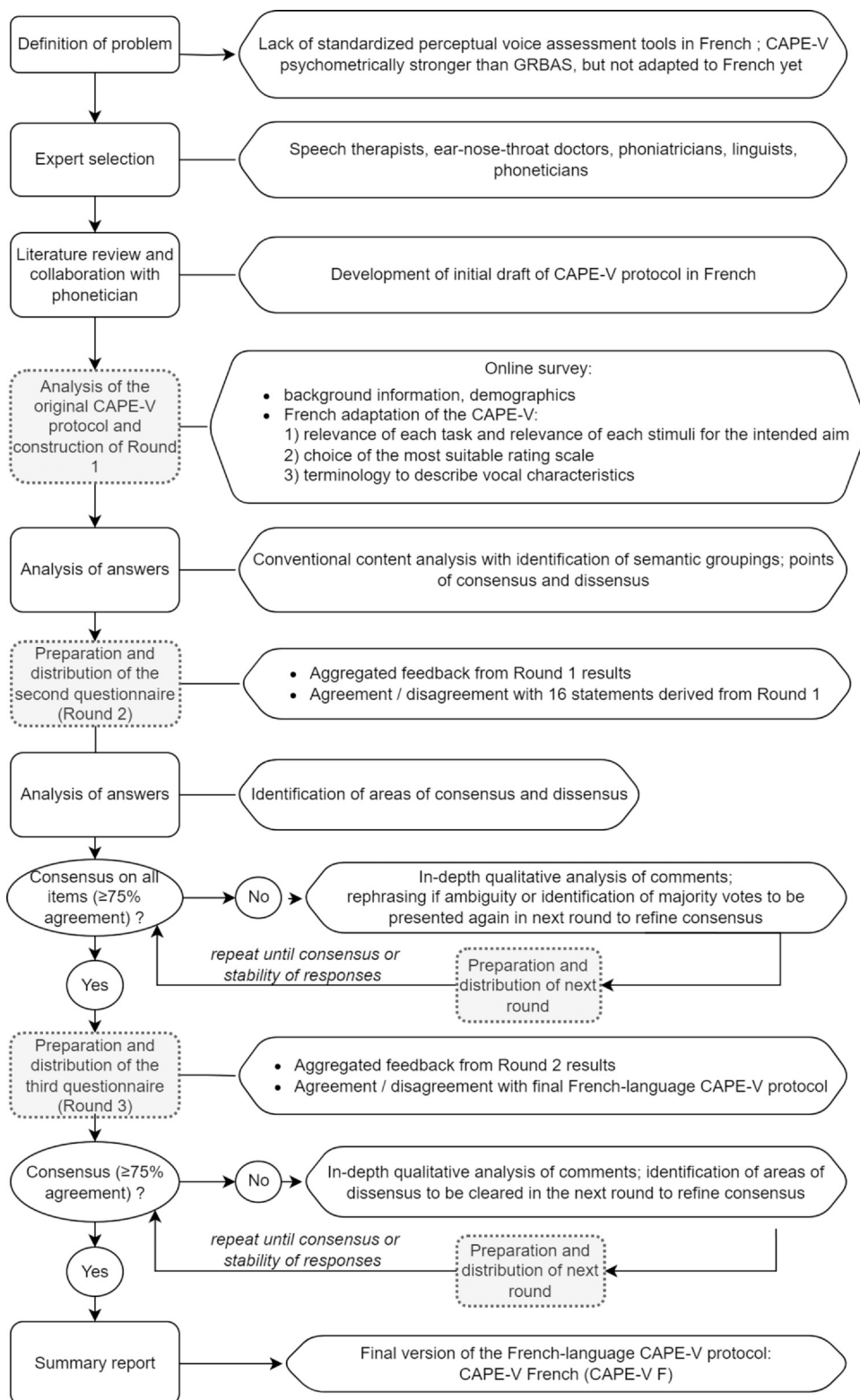


FIGURE 1. Flowchart of the Delphi process used in this study.

to reach a consensus on the adapted protocol. While this work has not been published nor validated and is not found in scientific databases, it is said to be used by clinicians in Quebec. We therefore contacted the head of the Community of Practice to learn about the history of the

Quebec-French adaptation of the CAPE-V and shared the progress of our research project with them. Subsequently, we incorporated certain elements from the Quebec version to build a protocol that also aligns with the criteria defined by this Community of Practice.

Delphi rounds and data analysis

This Delphi consensus survey was conducted over three consecutive rounds, between June and August 2023.

The survey questionnaires were distributed through a secure online platform hosted by the Faculty of Psychology, Speech and Language Therapy, and Education (UDI-Fapse). Each questionnaire was piloted by two researchers to get an estimate of the response time and to detect and correct potential execution problems (glitches and logical structure issues). Each round was available for 3 weeks, followed by a 2-week analysis and synthesis period during which the next survey was constructed.

In each round, participants who did not agree with a statement were encouraged to explain the reason for their disagreement, but comments were not mandatory so as not to bias toward positive answers.

The qualitative and quantitative data were analyzed using descriptive statistics (frequencies and percentages, means, standard deviations, medians, and interquartile ranges).

Round 1. The first round was composed of two parts.

In the first part, demographic data were collected (incl. occupation, years of experience with voice disorders, percentage of voice disordered patients, number of voice assessments per trimester, country and language of practice). After a description of the original CAPE-V protocol, the participants were also asked about the relevance of adapting the CAPE-V to French.

In the second part, the initial draft of the French adaptation was provided, with:

- Twenty mandatory single-choice questions about the content and the form of the protocol, targeting the relevance of each task of the protocol, the relevance of each stimulus inside of each task, the suitability of each stimulus with regards to its specified aim, and the suitability of different types of rating scales.
- Six mandatory open-ended (single-word answer) questions about the French terminology relating to different voice qualities; the participants were presented with voice samples (sustained vowel and continuous speech) selected for their prominent vocal quality. Six samples were presented, one from a female and one from a male speaker for each of the three qualities assessed in the original CAPE-V protocol (roughness, breathiness, strain).
- Twenty-seven single-choice and open-ended conditional questions depending on the responses to the mandatory answers.

Round 2. Responses and comments from round 1 were analyzed in an anonymized manner by the three authors. They were then synthesized and fed back to the participants for contextualization in round 2, together with 16 new questions (13 single-choice and 3 open-ended questions) based on the previous responses. The participants were first asked to choose and explain the most appropriate vowel(s) for the sustained vowel task among the most

frequent responses from round 1. They then indicated their degree of agreement (4-point Likert scale) with the six modified sentences for the sentence reading task. They were also asked to provide an open-ended answer with regards to their opinion on neutralizing the emotional loading of the question used to elicit semi-spontaneous speech. Three single-choice questions related to the terminology of vocal qualities based on responses from round 1. Two single-choice questions aimed at identifying the optimal rating scale for the pitch and loudness criteria, and finally, the participants responded to relevance of rating the pitch (binary answer with explanation).

Round 3. In the third round, the final version of the French CAPE-V protocol was presented to the expert panel. Each participant indicated their degree of agreement with the final protocol on a 4-point Likert scale and explained their agreement with an open answer.

Consensus and stop criteria

In the literature, consensus is primarily defined as the degree of agreement among participants, expressed as a percentage.³⁹ Consensus was considered defined a priori as the agreement of at least 75% of the participants.^{37,39} The planned limit of rounds was four, and the stop criterion was either an insufficient number of responses in two consecutive rounds (ie, less than 50% of participants), the obtention of the desired consensus, or stability in the responses.

RESULTS

Delphi panel

More than 200 individuals and institutions were directly contacted via email, and announcements were posted on 19 French-speaking speech therapy Facebook groups. Over 100 people clicked on the link leading to round 1. A total of 16 experts completed round 1—one participant was excluded as they did not have at least 5 years of experience in the field. A total dropout of 50% from rounds 1 to 3 was observed.

Detailed data for the participants in each round are available in [Table 1](#). All participants except one were speech therapists ($n = 14$), conducting an average of 12 voice consultations per week. Fifty-six percent of the participants had over 20 years of experience in the field of voice. The participants were primarily Belgian (53%) and French (47%). All participants considered the adaptation of the CAPE-V in French to be useful. The diversity of the panel was essentially the same throughout the rounds with regards to their occupation and country and language of practice, but the second and third rounds consisted of overall more experienced participants.

Stimuli to elicit voice samples

Sustained phonation

In our initial draft, we did not specify which vowel to use so as not to bias the panel's responses. The experts were asked to indicate and explain the most relevant vowel(s). During

Table 1.
Description of the Expert Panel

Round	1	2	3*
Participants (n)	15	9	8
<i>Occupation</i>	<i>N (%)</i>	<i>N (%)</i>	<i>N (%)</i>
Speech therapist only	9 (60%)	5 (56%)	4 (50%)
Speech therapist and lecturer	3 (20%)	2 (22%)	2 (25%)
Speech therapist and researcher	2 (13%)	1 (11%)	0 (0%)
Researcher and lecturer	1 (7%)	1 (11%)	1 (13%)
<i>Number of weekly voice consultations</i>			
Mean	12.14	17	19.5
STD	10.48	11.38	11.73
Min	2	2	2
Max	35	35	35
<i>Proportion of patients with voice disorders</i>	<i>N (%)</i>	<i>N (%)</i>	<i>N (%)</i>
Not applicable	1 (7%)	1 (11%)	1 (13%)
0%–25%	9 (60%)	4 (44%)	3 (38%)
25%–50%	2 (13%)	1 (11%)	0 (0%)
50%–75%	2 (13%)	2 (22%)	2 (25%)
75%–100%	1 (7%)	1 (11%)	1 (13%)
<i>Number of voice assessments per quarter</i>			
Not applicable	1 (7%)	1 (11%)	1 (13%)
0–2	7 (47%)	4 (44%)	2 (25%)
3–6	2 (13%)	2 (22%)	2 (25%)
7–9	1 (7%)	0 (0%)	0 (0%)
10–12	1 (7%)	0 (0%)	0 (0%)
13–15	0 (0%)	0 (0%)	0 (0%)
> 15	3 (20%)	2 (22%)	2 (25%)
<i>Country of practice</i>			
Belgium	8 (53%)	5 (56%)	3 (38%)
France	7 (47%)	4 (44%)	4 (50%)
<i>Language of practice</i>			
French	13 (87%)	7 (78%)	5 (63%)
French and English	1 (7%)	1 (11%)	1 (13%)
French and Créole French	1 (7%)	1 (11%)	1 (13%)
<i>Years of experience in the field of voice</i>			
5–10 years	3 (20%)	1 (11%)	2 (15%)
10–15 years	2 (13%)	1 (11%)	0 (0%)
15–20 years	1 (7%)	0 (0%)	0 (0%)
> 20 years	9 (60%)	7 (78%)	5 (63%)
<i>Self-attributed level of expertise</i>			
Novice	1 (7%)	1 (11%)	1 (13%)
Intermediate	8 (53%)	4 (44%)	3 (38%)
Expert	6 (40%)	4 (44%)	3 (38%)

Notes: The total percentages do not necessarily amount to 100% due to rounding.

* During round 3, the identity of one of the eight participants could not be confirmed because of a missing email address; it was considered as missing demographic data (13%).

the first round, all participants agreed that the sustained vowel task provides relevant information during perceptual voice evaluation. The spontaneously most mentioned vowels were /a/ (79%), /i/ (43%), and /u/ (21%).

Based on this observation, in the second round, the panel was asked to reach a consensus on which vowel(s) to retain.

Nearly half of the experts (56%) agreed to include only the vowel /a/. They argued that it is a relatively neutral vowel that limits the implication of the upper resonators and simultaneously allows for the assessment of maximum phonation time. Furthermore, this choice was related to their habit of using the vowel /a/ in voice assessments. Twenty-two percent of the panel argued that all three cardinal vowels should be included in the protocol because it allows for the investigation of different phonatory configurations and their impact on the patient's voice.

Based on the majority consensus, only the vowel /a/ was included in the final protocol submitted for approval in round 3.

Sentence reading

Sentence 1. The first sentence aims to analyze the production of the cardinal vowels (a, i, u). To do this, we initially proposed the following sentence: “Dis-nous Marie, où allez-vous samedi?” ([dinumaxi.ualevusamædi], “Tell us, Marie, where are you going on Saturday?”). Eighty-seven percent of the participants considered the analysis of the cardinal vowels to be relevant for the perceptual evaluation of voice. Among them, all considered that the sentence was suitable for the intended purpose, resulting in a consensus. Nevertheless, we considered the participants' comments to improve the sentence. They particularly mentioned that the sentence and the words used were quite short and that the sentence structure could imply a grammatical error (ie, addressing the subject “Marie” and then using the plural personal pronoun “vous”). One participant questioned the use of an interrogative sentence, which implies a different melodic pattern than the other sentences in the task. We chose to retain this variation as it provides a dynamic vocal sample, which is relevant within the context of a perceptual voice evaluation.

Taking these suggestions into account, we proposed the following sentence for the second round: “Dites-nous les amis, où allez-vous samedi?” ([ditnulezami.ualevusamædi], “Tell us, friends, where are you going on Saturday?”). All participants considered this sentence suitable for analyzing the production of cardinal vowels. A consensus was thus reached for this sentence, which was included in the final protocol.

Sentence 2. The second sentence is intended to analyze the transitions between voiced and voiceless phonemes. Unlike English, French does not use word-initial /h/-sounds. Hence, we initially proposed the following sentence: “Son chat Fanfreluche fonce sans sourciller.” ([sɔ̃ʃafɛ̃fɛ̃lyʃfɔ̃ssasurɔ̃sjɛ], “Their cat Fanfreluche goes for it without batting an eyelid”). However, only 67% of the participants considered this analysis relevant for the perceptual assessment of voice. From the provided rationales, it became evident that they were questioning the sentence itself rather than the targeted aim. This percentage, therefore, was taken with caution. We relied

on the specific feedback from participants to construct a more relevant sentence, as the proposed version did not reach satisfaction. Participants pointed out that the sentence was difficult to pronounce, such as a tongue twister. Moreover, the sentence contained many nasal vowels, thus risking bias with regards to velopharyngeal function.

We took these comments into account to improve the proposed sentence. In the second round, we presented them with the following sentence: “Son chat Fifi saute sur la souris.” ([sɔ̃ʃafifisotsyʁlasurɪ], “Their cat Fifi jumps on the mouse”). This simplified version ensured that the analysis was not biased by articulation. We also added oral vowels to facilitate production. Eighty-nine percent of participants considered the sentence suitable for the intended purpose. Consensus was reached for this sentence, and we have included it in the final protocol.

Sentence 3. The sentence “Elle vit aux Iles de la Madeleine” ([ɛlvi(t)ozildəlamadəlɛn], “She lives in the Magdalen Islands”) is fully voiced and aims to analyze the ability to maintain phonation throughout the whole utterance and to detect the presence of devoicing episodes or spasms. All participants reached a consensus: they considered the analysis relevant for perceptual voice evaluation, and the sentence was suitable for the objective.

However, based on a comment pointing out that the optional liaison of “vit_aux” could induce the production of a voiceless phoneme ([vito]), we proposed a modified sentence as follows in the second round: “Elle va aux Iles de la Madeleine” ([ɛlvaozildəlamadəlɛn], “She goes to the Magdalen Islands”). All participants approved of this last sentence, which was included in the final protocol.

Sentence 4. To assess hard glottal onsets, we initially proposed the sentence: “Irma, où est Alain ? Hier, il a atterri à Ottawa” ([iʁma.œ(t)alɛ̃/jɛʁila:tɛʁjaɔtawa], “Irma, where is Alain? Yesterday, he landed in Ottawa”), which contains several word-initial vowels. Eighty-seven percent of the panel considered this analysis relevant for the perceptual evaluation of voice; among them, 92% affirmed that the sentence was suitable for the objective. However, we shortened the sentence based on participants’ feedback describing it as too long. Furthermore, the word “hier” ([jɛʁ], “yesterday”), starting with a glide which would not allow to detect patients who produce a hard onset, was replaced with the word “aujourd’hui” ([oʒurdɥi], “today”). Lastly, each word preceding a hard-onset target should end with a vowel to avoid liaison (eg, avoid “est_Alain” [ɛ(t)alɛ̃], prefer “va Alain”).

Based on these comments, we proposed the following sentence in the second round: “Irma, où va Alain aujourd’hui? A Ottawa” ([iʁma.uva:lɛ̃oʒurdɥi/ɔtawa], “Irma, where is Alain going today? To Ottawa”). Eighty-nine percent of the panel approved of the proposed sentence, which was thus included in the final protocol.

Sentence 5. The sentence “Lundi, maman nous a montré les palmiers de Miami” ([lœ̃di.mamānuzamɔ̃tʁɛləpalmjedəmiʝami], “On Monday, Mom showed us the palm trees of Miami”) was designed to assess hyponasality. Ninety-three percent of the panel considered this objective relevant for perceptual voice evaluation. Among them, 93% stated that the sentence was suitable for the objective. Participants noted that the sentence was too long. Additionally, the sound [ɛ̃] was missing, which we incorporated with the French word “matin” (morning) in the new sentence for the second round (note that in English, unlike in French, there are no nasal vowels).

The following modified sentence was presented in round 2: “Lundi matin, maman nous a montré les nénuphars” ([lœ̃dimatɛ̃.mamānuzamɔ̃tʁɛlənenyfaʁ], “On Monday morning, Mum showed us the water lilies”). Participants unanimously agreed to validate the sentence, which was included in the final protocol.

Sentence 6. The sentence “Coco calcule tout pour payer tes trois pastèques” ([kokokalkyltupɛʁɛjetɛtɛkwapastɛk], “Coco calculates everything to pay for your three watermelons”) was constructed to assess nasal leakage and the impact of voiceless plosives on phonation. In the first round, 87% of the participants considered this analysis relevant in the perceptual voice evaluation. Among them, 92% affirmed that the sentence was suitable. However, the panel provided some feedback for improvement. First, the sentence was found to be too long. Second, the use of the phoneme [u] induces an excessive velar lowering. Finally, the participants wished to incorporate voiced plosives, suggesting that nasal leakage is better perceived on these phonemes.

The sentence was thus modified as follows: “Dobi comptait payer les deux pastèques” ([dobikɔ̃tɛʁɛjeledɔpastɛk], “Dobi intended to pay for the two watermelons”). This yielded a 78% agreement among the participants in the second round; hence this sentence was included in the final protocol.

Semi-spontaneous speech

In the original protocol, the semi-spontaneous speech is prompted by the question, “Tell me about your voice problem” or “Tell me how your voice is functioning.” These questions can elicit emotions in the patient and affect their voice. Therefore, we proposed neutralizing this emotional loading to obtain a more neutral sample of the patient’s voice. In the first round, 73% of the participants considered it relevant to elicit semi-spontaneous speech without emotional load. Among them, 82% affirmed that the question, “Aimez-vous les fruits? Pourquoi?” (“Do you like fruit? Why?”) was suitable for this purpose. However, participants questioned the use of the verb “aimer” (“like/love”) which they considered not neutral enough. Moreover, one participant did not agree with the interest of avoiding emotional loading.

Hence, in the second round, the participants were specifically asked to choose between neutralizing and retaining

the emotional loading. It was challenging to make a clear decision as several comments did not answer the question or take a position on either response option. Ultimately, the participants favored a question that neutralizes the emotional loading. In round 3, we suggested a different sentence than the previous one, as the latter was met with criticism: “Présentez-vous brièvement (nom, prénom, région, activité principale)” [“Briefly introduce yourself (last name, first name, region, main activity)”].

Rating scales

The original CAPE-V protocol uses a VAS with three categorical benchmarks to guide the rating. For the French adaptation, five types of rating scales were proposed to the panel in round 1 to assess their suitability for the perceptual assessment of voice: the semantic Likert scale (ordinal scale with labels, eg, “Strongly agree, Agree... Strongly disagree”, without numbers), the mixed Likert scale (ordinal scale with labels and numbers, eg, “1. Strongly agree, 2. Agree... 5. Strongly disagree”), the simple numerical scale (numbered scale without labels, eg, from 0 to 10), the notched VAS (a line with two labeled endpoints and notches/numerical or verbal markers on the line), and the simple VAS (a line with two labeled endpoints, with the rest of the line being empty of labels and values).

Only the simple received most favorable votes, as 67% of the participants consider it suitable for the perceptual evaluation of voice. They state that the continuum allows for a detailed analysis while avoiding the bias of categorization. The scale is easy to use, and the rating is precise. Based on the results, the simple VAS was included in the final protocol for rating vocal characteristics.

Rating scales: pitch and loudness

The rating modality for pitch and loudness were discussed separately, as the scales in the original protocol are slightly different. They can vary between three poles: too low/soft, normal, or too high/loud. Hence, in the original protocol, the rater first needs to specify the direction of the alteration with a key word (ie, “too high/loud” or “too low/soft”) and then indicate the degree of alteration on the VAS.

We questioned whether this was the most appropriate way to assess pitch and loudness. In the first round, we proposed three rating modalities to the panel: the original CAPE-V method, a qualitative verbal description of the altered criterion, or a bipolar labeled VAS with two directional extremes (eg, for loudness, “too soft” on the left, “too loud” on the right, with a normal loudness being the center of the scale). Participants preferentially chose the original CAPE-V method (47%) and the bipolar scale (40%). In round 2, we asked the panel to choose between these two options. Once again, a clear majority could not be obtained, as the responses were almost evenly divided, with a one-vote difference. Eventually, the panel (56%) opted for the VAS from the original protocol.

Terminology to describe vocal qualities

We consulted the panel of experts regarding the terms to use to describe the main vocal characteristics during the perceptual assessment of voice.

When presented with the voice samples with salient “roughness,” participants spontaneously primarily used the terms “rauque/raucité” (N = 13) and “érraillée/érraillement” (N = 6) in the first round. In the second round, they reached a consensus of 78% in using the term “rauque/raucité.”

For the samples predominantly presenting with “breathiness,” in the first round, participants spontaneously mostly used the terms “soufflée/souffle” (N = 19), “hypotonique/hypotonie” (N = 3), and “faible/faiblesse” (N = 2). In the second round, the panel reached a consensus of 89% in using the term “soufflée/souffle.”

When listening to samples with salient “strain,” participants spontaneously described it as “serrée/serrage” (N = 9) and “forcée/forçage” (N = 7) in the first round. In the second round, participants reached a consensus of 67% in using the term “serrée/serrage.” Although full consensus was not achieved, this was the term preferred by the majority of the panel, thus it was proposed in the final round for validation.

Final CAPE-V protocol in French

The final formatted protocol of the CAPE-V in French is given in [Appendix A](#).

DISCUSSION

This study aimed at adapting the CAPE-V protocol to the French language, based on the opinions of international voice experts. Consensus was sought regarding the content of the three protocol tasks, the rating scale to be used, and the terminology to describe the assessed vocal characteristics. The final protocol is relatively similar to the original English version, as it maintains the structure and core objectives. It includes a sustained vowel task, sentence reading, and semi-spontaneous speech. The voice is evaluated using a VAS for the same six criteria present in the original protocol: overall severity, roughness, breathiness, strain, pitch, and loudness, with the option to assess two additional criteria using two blank VASs.

Adaptations of the protocol into other languages have been carried out by small groups of phoneticians, linguists, or speech therapists (between one and four participants),^{16–29} mostly using non-systematic methods. In the present study, we report the results of a more systematic approach, involving a larger group of experts.

Delphi method and expert panel

The Delphi method was chosen for its many advantages over other methods. The consensus development conference, for example, also involves a panel of experts, who gather for a few days (including all-night sessions) to agree on a statement. It is rather tedious to organize and does not

consider issues related to group settings, such as social pressure to comply.³⁵ The nominal group and the Delphi methods, on the other hand, are more structured and systematic approaches, involving a multi-stage process.^{31,42} The Delphi method has the advantage of bypassing the face-to-face setting, thus allowing for larger panels and unlimited geographic reach, as well as for a control of the group's influence on responses via quasi-anonymity.^{38,41}

Although the sample size in Delphi studies can have a significant impact on consensus indexes,³³ it may range from six to more than 1000 experts, dependent on the topic and available resources.⁴³ Distributing the questionnaires online allowed us to reach more than 200 individuals and institutions from all French-speaking regions around the world. Our initial panel size of 15 experts corresponded to the numbers reported in numerous studies, ie, 12 to 15 experts.⁴⁴ While a dropout rate of 20% to 30% was expected,⁴⁰ we observed a higher dropout of 50%. Efforts were taken to minimize attrition to avoid an overestimation of the final consensus,⁴¹ by initially inviting participants to only take part if involvement in all the three expected rounds was envisaged. The resulting panel size remained larger than the groups involved in previous adaptations of the CAPE-V protocol, and its characteristics were representative of the main target users—French-speaking speech therapists with a substantial caseload of voice-disordered patients. Hence, overall, the Delphi panel was considered satisfying to address the main objective of this study.

Task stimuli

Throughout the iterative Delphi process, quantitative data were collected to objectify the consensus. Furthermore, qualitative data allowed for detailed feedback on the proposed adaptations, guiding the construction of the French CAPE-V protocol and maximizing the probability of this tool to be adopted by voice professionals. The use of both quantitative and qualitative data is challenging yet more informative. Quantitative majority was favored over qualitative feedback to respect the systematic nature of our approach. When the recommended agreement threshold of 75% of the panel³⁹ was not reached, qualitative feedback allowed us to fine-tune the statements for the subsequent round, until consensus was reached.

For the sustained vowel task, the vowel /a/ was favored both in rounds 1 and 2. The participants' comments were mixed, some of them providing strong arguments (eg, a relatively neutral vowel that limits the implication of the upper resonators), others seeming more reluctant: "Out of habit, I suppose," "I would have included /a/ and /u/," "If we want to reach a consensus, only /a/ is fine, that's what is done in practice". In contrast, the 22% of participants who chose the combination of the vowels /a/, /i/, and /u/ defended their choice with more concrete elements, eg, "The three extremes of the vowel triangle: /i/ is the tensest, /u/ is the most relaxed, and /a/ reflects the balance between the two," "anterior/posterior position and openness/closure in

the vowel triangle." Despite the quality of these comments, the quantitative majority was favored. Hence, the single vowel /a/ option was presented in the last round and was approved by all experts.

The consensus achieved for the sentences was substantial and unambiguous. The adaptation of the sentences represented the core of the translation work of the protocol, as the other tasks did not require translation per se, but rather a consensus on the choice of the stimulus. The comments from the participants regarding the suggested sentences were very constructive and allowed to reach high consensus levels as well as refinement of the sentences even when consensus was reached. Indeed, in the second round, half of the sentences yielded a perfect consensus, while the other half obtained a consensus of over 75%. The final set of sentences was unanimously approved.

For the semi-spontaneous speech task, the experts agreed on the use of a question without emotional load. However, although the consensus threshold was also reached for the relevance of the provided question ("Do you like fruits? Why?"), nearly a third of the responses were ambiguous, such as "Both answers suit me," "I would have asked about a mundane question. And a more intimate question," "Using the verb 'like' doesn't bother me," and "No, the emotional charge is low when talking about a taste. [...] I don't see a powerful commitment in it. It's like 'do you prefer rain or sunshine'." Hence, although the neutralization of the emotional load seemed to be favored, the question itself was modified for the last round. The new question ("Briefly introduce yourself [last name, first name, region, main activity]") was approved by the entire panel.

Rating scale and voice terminology

The choice of the rating scale ultimately settled on a simple VAS for all vocal characteristics. From the very first round, participants disapproved of all rating scales except for the VAS. Although the threshold of 75% was not reached, the panel's disapproval with the other options was adamant. Hence, the quantitatively favored option was once again selected and approved by all experts in the final protocol. Of note, the VAS has been described in the literature as a reliable⁴⁵ and well-accepted method to measure the perception of various clinical phenomena,^{46,47} including in the field of voice.^{48–50} It has been shown that the VAS allows for finer ratings of voice quality as compared to ordinal scales.⁵¹ Furthermore, including numbers or verbal descriptors at intermediate points along the line (such as those present on the original English CAPE-V form) has been advised against, as this could lead to clustering of points around specific numerical values.^{48,52,53}

The agreement on terms used to describe voice qualities is a first step toward a consensus in the French-speaking community. In the first round, for each vocal characteristic, two or three terms stood out, indicating a shared vocabulary among voice experts. We regret the absence of participants from Quebec, Switzerland, and North Africa, who would have brought additional input to this consensus.

Indeed, the unpublished Quebec version of the CAPE-V protocol, for instance, uses different terms than those adopted at the end of the present Delphi process, such as “éteinte” (smothered) for “breathiness,” while French-speaking experts from Belgium and France favor the word “soufflée” (breathy).

LIMITS AND PERSPECTIVES

While our study achieved a larger sample size compared to previous translation efforts of the CAPE-V protocol, it is essential to acknowledge certain limitations and consider them in a constructive light. Despite the considerable number of targeted French-speaking stakeholders, the participation rate may have been influenced by certain factors. One potential deterrent may have been the explicit request for participants to commit to multiple rounds within a 3-month timeframe on the first page of the initial questionnaire. This may have discouraged many of the more than 100 participants who accessed the link. Additionally, the timing of our survey launch, between June and August, coincided with a period of high workload for many researchers, lecturers, and clinicians at the end of the academic year, followed by the holiday season. Recognizing this observation is crucial for future surveys using the Delphi method, allowing for more strategic planning.

Another limitation we acknowledge is the lack of diversity in our participant panel, consisting exclusively of individuals from France and Belgium, without representation from Switzerland, Quebec, or North African French speakers. The use of the French protocol is thereby to be used with caution in the latter regions, as cultural differences and dialects can have a significant impact on the sentence reading task. The panel, primarily comprising speech therapists and lecturers, missed the inclusion of professionals, such as ear, nose, and throat doctors, phoniatricians, and linguists. While this narrowed our focus to the needs and expectations of French and Belgian speech therapists, it limits the generalizability of our protocol. Despite the difficulties faced by researchers today to engage experts in their studies,^{54–57} our panel demonstrated the high interest of clinicians, dealing with patients on a daily basis, in improving the set of available voice assessment tools. It highlights the dedication of professionals to contribute to improvements in their field.

Achieving the required consensus on all questionnaire items proved challenging—a common phenomenon according to Niederberger and Spranger’s meta-analysis.⁵⁸ The authors advocate for a flexible approach to consensus definition while maintaining methodological coherence. Hence, we opted for a majority agreement (> 50%) for the rating scale, deviating from our original methodology. This decision was supported by the participants’ comments and explicit disapproval with other rating methods, as well as by existing literature emphasizing the relevance of using simple VAS.

Despite these limitations, the results obtained in the final round yielded a remarkable 100% consensus, underscoring the relevance and robustness of the final protocol. A first validation study of the new protocol in French is being conducted at the University of Toulouse (France) on both dysphonic and non-dysphonic populations, with the GRBAS and the Voice Handicap Index⁵⁹ as concurrent outcome measures. More than 50 French clinicians have already expressed their interest in this tool. Further validation in various French-speaking regions around the world (France, Belgium, Switzerland, Quebec, and Maghreb) is essential for making the CAPE-V widely accessible to French-speaking clinicians seeking standardized and validated voice assessment tools.

CONCLUSION

This work combined the benefits of quantitative and qualitative analyses to propose a relevant adaptation of the CAPE-V protocol for perceptual voice assessment in French. The Delphi process, involving a panel of Belgian and French voice experts, considered both the percentages of agreement and the constructive comments of the panel members in an iterative and collaborative three-round process. The proposed version maintains the original goals of the English protocol, while offering relevant and adapted supports for use in vocal assessment contexts in Belgium or France. Preliminarily validated by the consulted panel of experts, the use of this new protocol in clinical and research settings now awaits further validation.

Declaration of Competing Interest

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

Acknowledgments

The authors extend their sincere appreciation to Professor Laurent Rasier, Associate Professor, in the Department of Languages and Modern Literature at the University of Liège, whose expertise in the French language played a fundamental role in enriching the quality and precision of our adaptation process. The authors also express their gratitude to the voice experts whose unwavering participation and valuable insights were integral to the realization of this study. Their active engagement in the research, providing constructive feedback and expertise, played a vital role in achieving consensus for the adaptation of the CAPE-V protocol into the French language.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.jvoice.2024.03.011](https://doi.org/10.1016/j.jvoice.2024.03.011).

References

- Lechien JR, Geneid A, Bohlender JE, et al. Consensus for voice quality assessment in clinical practice: guidelines of the European Laryngological Society and Union of the European Phoniaticians. *Eur Arch Otorhinolaryngol*. 2023;280:5459–5473. <https://doi.org/10.1007/s00405-023-08211-6>.
- Stachler RJ, Francis DO, Schwartz SR, et al. Clinical practice guideline: hoarseness (dysphonia) (update). *Otolaryngol Head Neck Surg*. 2018;158(suppl 1):S1–S42. <https://doi.org/10.1177/0194599817751030>.
- Yamasaki R, Madazio G, Leão SHS, et al. Auditory-perceptual evaluation of normal and dysphonic voices using the voice deviation scale. *J Voice*. 2017;31:67–71. <https://doi.org/10.1016/j.jvoice.2016.01.004>.
- Lee Y, Kim G, Kwon S. The usefulness of auditory perceptual assessment and acoustic analysis for classifying the voice severity. *J Voice*. 2020;34:884–893. <https://doi.org/10.1016/j.jvoice.2019.04.013>.
- Oates J. Auditory-perceptual evaluation of disordered voice quality. *Folia Phoniatr Logop*. 2009;61:49–56. <https://doi.org/10.1159/000200768>.
- Ghio A, Revis J, Smithson-Barrière D, et al. Reliability and correlations between overall severity, roughness and breathiness in the perception of dysphonic voices: investigating cognitive aspects. *J Voice*. 2021;38:136–143. <https://doi.org/10.1016/j.jvoice.2021.07.010>.
- Hirano M. *Clinical Examination of Voice*. Wien, Germany: Springer Verlag; 1981.
- Nemr K, Simões-Zenari M, Cordeiro GF, et al. GRBAS and Cape-V scales: high reliability and consensus when applied at different times. *J Voice*. 2012;26:812.e17–812.e22. <https://doi.org/10.1016/j.jvoice.2012.03.005>.
- 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*. 2021. <https://doi.org/10.1016/j.jvoice.2021.10.022>.
- Wang Z, Yu P, Yan N, et al. Automatic assessment of pathological voice quality using multidimensional acoustic analysis based on the GRBAS scale. *J Signal Process Syst*. 2016;82:241–251. <https://doi.org/10.1007/s11265-015-1016-2>.
- Karnell MP, Melton SD, Childes JM, et al. Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *J Voice*. 2007;21:576–590. <https://doi.org/10.1016/j.jvoice.2006.05.001>.
- 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. [https://doi.org/10.1044/1058-0360\(2008/08-0017\)](https://doi.org/10.1044/1058-0360(2008/08-0017)).
- Fujiki RB, Thibeault SL. The relationship between auditory-perceptual rating scales and objective voice measures in children with voice disorders. *Am J Speech Lang Pathol*. 2021;30:228–238. https://doi.org/10.1044/2020_AJSLP-20-00188.
- 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\(2010/09-0105\)](https://doi.org/10.1044/1058-0360(2010/09-0105)).
- Kelchner LN, Brehm SB, Weinrich B, et al. Perceptual evaluation of severe pediatric voice disorders: rater reliability using the consensus auditory perceptual evaluation of voice. *J Voice*. 2010;24:441–449. <https://doi.org/10.1016/j.jvoice.2008.09.004>.
- Mohd Mossadeq N, Mohd Khairuddin KA, Zakaria MN. Cross-cultural adaptation of the Consensus Auditory-perceptual Evaluation of Voice (CAPE-V) into Malay: a validity study. *J Voice*. 2022;S0892-1997:00151-5. <https://doi.org/10.1016/j.jvoice.2022.05.018>.
- Núñez-Batalla F, Morato-Galán M, García-López I, et al. Validation of the Spanish adaptation of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *Acta Otorrinolaringol (Engl Ed)*. 2015;66:249–257. <https://doi.org/10.1016/j.otoeng.2015.08.001>.
- Joshi A, Baheti I, Angadi V. Cultural and linguistic adaptation of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) into Hindi. *J Speech Lang Hear Res*. 2020;63:3974–3981. https://doi.org/10.1044/2020_JSLHR-20-00348.
- de Almeida SC, Mendes AP, Kempster GB. The Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) psychometric characteristics: II European Portuguese version (II EP CAPE-V). *J Voice*. 2019;33:582.e5–582.e13. <https://doi.org/10.1016/j.jvoice.2018.02.013>.
- Ertan-Schlüter E, Demirhan E, Ünsal EM, et al. The Turkish version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V): a reliability and validity study. *J Voice*. 2020;34:965.e13–965.e22. <https://doi.org/10.1016/j.jvoice.2019.05.014>.
- Kondo K, Mizuta M, Kawai Y, et al. Development and validation of the Japanese version of the Consensus Auditory-Perceptual Evaluation of Voice. *J Speech Lang Hear Res*. 2021;64:4754–4761. https://doi.org/10.1044/2021_JSLHR-21-00269.
- Chen Z, Fang R, Zhang Y, et al. The Mandarin version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) and its reliability. *J Speech Lang Hear Res*. 2018;61:2451–2457. https://doi.org/10.1044/2018_JSLHR-S-17-0386.
- Behlau M, Rocha B, Englert M, et al. Validation of the Brazilian Portuguese CAPE-V instrument—Br CAPE-V for auditory-perceptual analysis. *J Voice*. 2022;36:586.e15–586.e20. <https://doi.org/10.1016/j.jvoice.2020.07.007>.
- Mozzanica F, Ginocchio D, Borghi E, et al. Reliability and validity of the Italian version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *Folia Phoniatr Logop*. 2013;65:257–265. <https://doi.org/10.1159/000356479>.
- Salary MN, Khoddami S, Drinnan M, et al. Validity and rater reliability of Persian version of the Consensus Auditory Perceptual Evaluation of Voice. *Aud Vestib Res*. 2017;23:65–74.
- Gunjawate DR, Ravi R, Bhagavan S. Reliability and validity of the Kannada version of the Consensus Auditory-Perceptual Evaluation of Voice. *J Speech Lang Hear Res*. 2020;63:385–392. https://doi.org/10.1044/2019_JSLHR-19-00020.
- Özcebe E, Aydınli FE, Tığrak TK, et al. Reliability and validity of the Turkish version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *J Voice*. 2019;33:382.e1–382.e10. <https://doi.org/10.1016/j.jvoice.2017.11.013>.
- Venkatraman Y, Mahalingam S, Boominathan P. Development and validation of sentences in Tamil for psychoacoustic evaluation of voice using the Consensus Auditory-Perceptual Evaluation of Voice. *J Speech Lang Hear Res*. 2022;65:4539–4556. https://doi.org/10.1044/2022_JSLHR-22-00169.
- Dabirmoghaddam P, Khoramshahi H, Dehqan A, et al. Construct and discriminant validity of the Persian version of the Consensus Auditory Perceptual Evaluation of Voice (CAPE-V). *J Voice*. 2022;36:876.e9–876.e15. <https://doi.org/10.1016/j.jvoice.2020.09.023>.
- Narea-Veas MS, Farias PG, Vázquez Fernández P. Consensus Auditory Perceptual Evaluation of Voice (CAPE-V): revisión sistemática de los métodos utilizados para su adaptación y validación. *Rev Investig Innov Cienc Salud*. 2023;5:178–204. <https://doi.org/10.46634/riics.206>.
- McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm*. 2016;38:655–662. <https://doi.org/10.1007/s11096-016-0257-x>.
- Romero-Collado A. Elementos esenciales para elaborar un estudio con el método (e)Delphi. *Enferm Intensiva*. 2021;32:100–104. <https://doi.org/10.1016/j.enfi.2020.09.001>.
- Birko S, Dove ES, Özdemir V, et al. Evaluation of nine consensus indices in Delphi foresight research and their dependency on delphi survey characteristics: a simulation study and debate on Delphi design and interpretation. *PLoS One*. 2015;10:1–14. <https://doi.org/10.1371/journal.pone.0135162>.
- Linstone HA, Turoff M. *The Delphi Method: Techniques and Applications*. Turoff M, Linstone HA, eds. New Jersey Institute of Technology; Newark, NJ, 2002.
- Letrilliart L, Vanmeerbeek M. À la recherche du consensus: quelle méthode utiliser? *Exercer*. 2011;22:170–177.
- Cunningham BJ, Kwok E, Turkstra L, et al. Establishing consensus among community clinicians on how to categorize and define

- preschoolers' speech and language impairments at assessment. *J Commun Disord.* 2019;82:105925. <https://doi.org/10.1016/j.jcomdis.2019.105925>.
37. Denman D, Kim JH, Munro N, et al. Describing language assessments for school-aged children: a Delphi study. *Int J Speech Lang Pathol.* 2019;21:602–612. <https://doi.org/10.1080/17549507.2018.1552716>.
 38. von der Gracht HA. Consensus measurement in Delphi studies. Review and implications for future quality assurance. *Technol Forecast Soc Change.* 2012;79:1525–1536. <https://doi.org/10.1016/j.techfore.2012.04.013>.
 39. Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol.* 2014;67:401–409. <https://doi.org/10.1016/j.jclinepi.2013.12.002>.
 40. Chalmers J, Armour M. The Delphi technique. In: Liamputtong P, ed. *Handbook of Research Methods in Health Social Sciences.* Singapore: Springer; 2019:715–735. https://doi.org/10.1007/978-981-10-5251-4_99.
 41. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS Med.* 2011;8:e1000393. <https://doi.org/10.1371/journal.pmed.1000393>.
 42. McMillan SS, Kelly F, Sav A, et al. Using the nominal group technique: how to analyse across multiple groups. *Health Serv Outcomes Res Methodol.* 2014;14:92–108. <https://doi.org/10.1007/s10742-014-0121-1>.
 43. Powell C. The Delphi technique: myths and realities. *J Adv Nurs.* 2003;41:376–382. <https://doi.org/10.1046/j.1365-2648.2003.02537.x>.
 44. McPherson S, Reese C, Wendler MC. Methodology update: Delphi studies. *Nurs Res.* 2018;67:404–410. <https://doi.org/10.1097/NNR.0000000000000297>.
 45. Safikhani S, Gries KS, Trudeau JJ, et al. Response scale selection in adult pain measures: results from a literature review. *J Patient Rep Outcomes.* 2018;2:40. <https://doi.org/10.1186/s41687-018-0053-6>.
 46. Byrom B, Elash CA, Eremenco S, et al. Measurement comparability of electronic and paper administration of visual analogue scales: a review of published studies. *Ther Innov Regul Sci.* 2022;56:394–404. <https://doi.org/10.1007/s43441-022-00376-2>.
 47. Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. *Res Nurs Health.* 1990;13:227–236. <https://doi.org/10.1002/nur.4770130405>.
 48. Martins PC, Couto TE, Gama ACC. Auditory-perceptual evaluation of the degree of vocal deviation: correlation between the Visual Analogue Scale and Numerical Scale. *Codas.* 2015;27:279–284. <https://doi.org/10.1590/2317-1782/20152014167>.
 49. San Segundo E, Skarnitzl R. A computer-based tool for the assessment of voice quality through Visual Analogue Scales: VAS-simplified vocal profile analysis. *J Voice.* 2021;35:497.e9–497.e21. <https://doi.org/10.1016/j.jvoice.2019.10.007>.
 50. Naunheim MR, Dai JB, Rubinstein BJ, et al. A visual analog scale for patient-reported voice outcomes: the VAS voice. *Laryngoscope Invest Otolaryngol.* 2020;5:90–95. <https://doi.org/10.1002/liv.2.333>.
 51. Wuyts FL, De Bodt MS, Van de Heyning PH. Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J Voice.* 1999;13:508–517. [https://doi.org/10.1016/S0892-1997\(99\)80006-X](https://doi.org/10.1016/S0892-1997(99)80006-X).
 52. Hawker GA, Mian S, Kendzerska T, et al. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res (Hoboken).* 2011;63(suppl 11)<https://doi.org/10.1002/acr.20543>.
 53. Rouve S, Didier A, Demoly P, et al. Numeric score and visual analog scale in assessing seasonal allergic rhinitis severity. *Rhinol J.* 2010;48:285–291. <https://doi.org/10.4193/Rhino09.208>.
 54. Vas A, D'sa P, Daud H, Kulkarni A, et al. Perceived barriers to participation in clinical research amongst trauma and orthopaedic community: a survey of 148 consultants and junior doctors in Wales. *Cureus.* 2021;13:e19694. <https://doi.org/10.7759/cureus.19694>.
 55. Ciemins EL, Mollis BL, Brant JM, et al. Clinician engagement in research as a path toward the learning health system: a regional survey across the northwestern United States. *Health Serv Manage Res.* 2020;33:33–42. <https://doi.org/10.1177/0951484819858830>.
 56. Williams J, Craig TJ, Robson D. Barriers and facilitators of clinician and researcher collaborations: a qualitative study. *BMC Health Serv Res.* 2020;20:1126. <https://doi.org/10.1186/s12913-020-05978-w>.
 57. Goldstein KM, Gierisch JM, Tucker M, et al. Options for meaningful engagement in clinical research for busy frontline clinicians. *J Gen Intern Med.* 2021;36:2100–2104. <https://doi.org/10.1007/s11606-020-06587-3>.
 58. Niederberger M, Köberich S. Coming to consensus: the Delphi technique. *Eur J Cardiovasc Nurs.* 2021;20:692–695. <https://doi.org/10.1093/eurjcn/zvab059>.
 59. Jacobson BH, Johnson A, Grywalski C, et al. The Voice Handicap Index (VHI). *Am J Speech Lang Pathol.* 1997;6:66–70. <https://doi.org/10.1044/1058-0360.0603.66>.