

Evidence of response process validity of a spectrographic voice analysis protocol

Evidência de validade baseada nos processos de resposta de um protocolo de análise espectrográfica da voz

Allan Carlos França da Silva¹ , Emmanuel Gustavo Rodrigues Diniz¹ , Maxsuel Avelino Alves de Paiva¹ , Itacely Marinho da Silva¹ , Saulo Iordan do Nascimento Silva¹ , Luiz Medeiros Araujo Lima Filho¹ , Leonardo Wanderley Lopes¹ 

ABSTRACT

Purpose: To develop the validity step based on the response processes of the Spectrographic Analysis Protocol (SAP). **Methods:** 10 speech therapists and 10 undergraduate students of the Speech Therapy course were recruited, who applied the SAP in 10 spectrograms, performed the evaluation of the PAE items, and participated in a cognitive interview (CI). The SAP was reanalyzed to reformulate or exclude items based on the responses. The chi-square test and the accuracy values were used to analyze the answers to the questionnaires and qualitative analysis of the CI data. **Results:** the participants achieved accuracy > 70% in most items of the SAP. Only seven items achieved accuracy ≤ 70%. There was a difference between presence vs. absence of difficulty in identifying items in the spectrogram. Most participants had no problem identifying the SAP items. In the CI, only six items did not correctly identify the intention, verified in the qualitative analysis. In addition, participants suggested excluding five items. **Conclusion:** After the validation step based on the response processes, the SAP is reformulated. Seven items were deleted, and two items were reformulated. Thus, the final version of the SAP after this stage was reduced from 25 to 18 items, distributed in the five domains.

Keywords: Voice; Voice disorders; Speech acoustics; Voice quality; Validation study

RESUMO

Objetivo: desenvolver a etapa de validade baseada nos processos de resposta do Protocolo de Análise Espectrográfica da Voz (PAEV). **Métodos:** foram recrutados dez fonoaudiólogos e dez alunos de graduação em Fonoaudiologia, que aplicaram o PAEV em dez espectrogramas, realizaram o julgamento dos itens do PAEV e participaram de uma entrevista cognitiva. A partir das respostas, o PAEV foi reanalisado para reformulação ou para exclusão de itens. Utilizou-se o teste Qui-Quadrado e os valores de acurácia para análise das respostas dos questionários, assim como análise qualitativa dos dados da entrevista cognitiva. **Resultados:** os participantes obtiveram acurácia maior que 70% na maioria dos itens do PAE. Apenas sete itens alcançaram acurácia menor ou igual a 70%. Houve diferença entre as respostas de presença versus ausência de dificuldade na identificação dos itens no espectrograma. A maioria dos participantes não teve dificuldade na identificação dos itens do PAEV. Na entrevista cognitiva, apenas seis itens não obtiveram correta identificação da intenção, conforme verificado na análise qualitativa. Além disso, os participantes sugeriram exclusão de cinco itens. **Conclusão:** após a etapa de validação baseada nos processos de resposta, o PAEV foi reformulado. Sete itens foram excluídos e dois itens foram reformulados. Dessa forma, a versão final do PAEV após essa etapa foi reduzida de 25 para 18 itens, distribuídos nos cinco domínios.

Palavras-chave: Voz; Distúrbios de voz; Acústica da fala; Qualidade de voz; Estudo de validação

Study carried out at Programa de Pós-Graduação em Modelos de Decisão e Saúde, Universidade Federal da Paraíba – UFPB – João Pessoa (PB), Brasil.

¹Universidade Federal da Paraíba – UFPB – João Pessoa (PB), Brasil.

Conflict of interests: No.

Authors' contribution: ACFS participated in the idealization, collection, data analysis and writing of the manuscript; EGRD participated in the idealization, collection, data analysis and writing of the manuscript; MAAP participated in the collection, data analysis and writing of the manuscript; IMS participated in the collection, data analysis and writing of the manuscript; SINS participated in the collection, data analysis and writing of the manuscript; LMALF participated in the analysis, interpretation and revision of the manuscript; LWL participated in the idealization, analysis, interpretation, writing and revision of the manuscript.

Funding: None.

Corresponding author: Leonardo Wanderley Lopes. E-mail: lwlopes@hotmail.com

Received: July 04, 2023; **Accepted:** November 30, 2023

INTRODUCTION

Acoustic analysis is part of the multidimensional assessment of the voice and provides complementary information to assess, diagnose, and monitor voice disorders. In speech-language-hearing (SLH) routine, acoustic analysis usefully evaluates individuals with vocal complaints by extracting measures that quantify and qualify characteristics of the vocal signal, such as jitter, shimmer, and glottal-to-noise excitation (GNE)⁽¹⁻³⁾. It is considered a reference for vocal assessment in clinical use and research, as it compares deviated vocal signals with normative data, helping understand vocal production with a highly reproducible noninvasive method^(2,3).

Most acoustic analysis methods consist of quantitative measurements extracted with highly reproducible algorithms, which compare and classify a vocal signal according to reference values^(2,4). However, this approach may be influenced by the sound pressure level during voice recording and the degree of aperiodicity in the signal. Moreover, the various pieces of software use different algorithms, which may compromise the clinical application and not reflect the phenomenon evaluated⁽⁵⁻⁷⁾.

Voice spectrography is the main method of qualitative acoustic analysis of the vocal signal. The spectrogram is a three-dimensional graph approached through descriptive evaluation to observe harmonics behavior, energy distribution according to frequency bands, and so forth. The main advantage of spectrography is that it analyzes different vocal signals, regardless of the degree of aperiodicity and noise in the emission⁽³⁾.

On the other hand, spectrogram acoustic inspection is mainly criticized for its subjectivity, as it depends on the evaluator's experience and specific training^(8,9). One of the ways to improve the quality of spectrographic analysis is to develop standardized protocols to train new clinicians. The available literature^(5,10) has proposals to classify and characterize the narrowband spectrogram – which, however, did not go through a validation process for use in a clinical context.

Hence, the Spectrographic Voice Analysis Protocol (SAP) began being developed to classify through spectrography individuals with and without vocal deviations⁽¹¹⁾. The authors used as a basis the recommendations⁽¹²⁾ to develop and validate instruments. SAP went through the first stage of the validation process to verify the evidence of content, clarity, and relevance of the items in each domain⁽¹³⁾.

SAP was created to analyze narrowband spectrograms of a sustained vowel. Its items can analyze all vowels, as long as they are emitted in a sustained manner and the clinician always uses the same vowel to enable inter and intrasubject comparisons. The vowel chosen for this study was [ε] because it is the most commonly used in vocal assessment in Brazil and it is an open, unrounded oral vowel with a more neutral and intermediate position in the vocal tract for Brazilian Portuguese⁽¹³⁾.

SAP currently has the following five domains: the beginning of the emission, temporal aspects of the emission, energy distribution in the trace, description of harmonics, and distribution of noise in the trace (Figure 1). To use SAP, the clinician must first visually inspect the spectrographic trace of the sustained vowel.

1) Beginning of the emission		
Item		
1	()	Presence of noise at the beginning of the emission
2	()	Presence of dark well-defined vertical striations at the beginning of the emission
3	()	High number of irregular harmonics at the beginning of the emission
4	()	No irregularity or observable noise at the beginning of the trace

2) Temporal aspects of the emission		
Item		
5	()	Presence of irregularities in the trace
6	()	Gradual loss of trace definition/energy
7	()	Presence of abrupt interruptions in the trace
8	()	Presence of regular spectrographic trace

3) Energy distribution in the trace		
Item		
9	()	Presence of harmonics with little brightness
10	()	Decreased energy and number of harmonics above 4000 Hz
11	()	Presence of harmonics above 4000 Hz
12	()	Decreased energy or number of harmonics up to 4000 Hz
13	()	Increased energy between 1000 and 3000 Hz
14	()	Predominance of F0 trace rather than that of other harmonics
15	()	Decreased energy level in the whole frequency band throughout the trace
16	()	Increased energy level in the whole frequency band throughout the trace
17	()	Series rich in regular harmonics up to 4000 Hz

4) Description of the harmonics		
Item		
18	()	Presence of irregular horizontal striations between harmonics
19	()	Presence of undefined or sketched harmonics
20	()	Predominant presence of low-amplitude harmonics
21	()	Presence of irregular harmonics

5) Distribution of noise in the trace		
Item		
22	()	Presence of noise between harmonics below 4000 Hz
23	()	Presence of additional diffuse noise above 4000 Hz
24	()	Harmonics replaced with noise in the spectrographic trace
25	()	Spectrographic trace with discrete or no noise between harmonics below 4000 Hz

1) Beginning of the emission		
Item		
a)	()	Presence of noise or irregularities at the beginning of the emission

2) Temporal aspects of the emission		
Item		
b)	()	Changes in the spectrographic trace configuration in the time domain
c)	()	Gradual loss of trace definition/energy
d)	()	Presence of abrupt interruptions in the trace

3) Energy distribution in the trace		
Item		
e)	()	Presence of harmonics with little brightness
f)	()	Decreased energy and number of harmonics above 4000 Hz
g)	()	Presence of harmonics above 4000 Hz
h)	()	Decreased energy or number of harmonics up to 4000 Hz
i)	()	Increased energy between 1000 and 3000 Hz
j)	()	Decreased energy level in the whole frequency band throughout the trace
k)	()	Increased energy level in the whole frequency band throughout the trace

4) Description of the harmonics		
Item		
l)	()	Presence of irregular horizontal striations between harmonics
m)	()	Presence of undefined or sketched harmonics
n)	()	Predominant presence of low-amplitude harmonics
o)	()	Presence of harmonics with irregular trajectory and morphology (non-rectilinear)

5) Distribution of noise in the trace		
Item		
p)	()	Presence of noise between harmonics below 4000 Hz
q)	()	Presence of additional diffuse noise above 4000 Hz
r)	()	Harmonics replaced with noise in the spectrographic trace

Figure 1. Spectrographic Voice Analysis Protocol: (A) version before the response process validity stage; (B) version after the response process validity stage

Next, they mark in SAP the items observed in the spectrogram. No score or cutoff point has yet been established for the instrument, as the SAP is in the validation process and is still subject to changes in psychometric and structural properties until validation is completed. Currently, SAP has been used to train clinicians for spectrographic descriptions of voice samples obtained from dysphonic patients. A score is expected to be defined at the end of the validation stages so that professionals can classify the presence, absence, and degree of vocal deviation based on the cutoff point.

Considering the sequence in an instrument validation process, this research aimed to address the SAP response process validity. This stage will investigate and analyze difficulties in using each SAP item and verify their comprehensibility so that it can be reformulated based on the results.

Response processes result from observations or judgments about the behavior or performance of different strata of the target population during the application of the test. At this stage, we sought to understand the psychological, cognitive, and social processes involved in applying the test and verify the adequacy, structure, and application of the items in a real context⁽¹⁴⁾.

METHODS

Study design

This instrument validation research was evaluated and approved by the Research Ethics Committee of the Federal University of Paraíba under evaluation report number 508.200/2013. All participants signed an informed consent form. The methodology used for this validation stage was based on recommendations available in the current literature^(12,15,16).

Participants

No explicit sample size recommendation is available for studies addressing response process validity. When this

validation stage involves cognitive interviews (CI), as is the case in this research, a sample of five to 15 participants representing each stratum of the target population is suggested. The exploratory nature of CI makes it more likely to detect problems in the instrument items in the response process validity stage with a smaller sample, as it reduces the variability of CI results^(17,18).

Thus, a convenience sample was recruited with 10 SLH pathologists and 10 undergraduate SLH students, who had already completed spectrographic analysis training in the required subject in the program. All participants were affiliated to, enrolled in, or graduated from the institution where this research was carried out, so they were trained in spectrographic analysis as required by their training, in the second year of the undergraduate program. The Fehring Model scoring system was adapted to select SLH pathologists to participate in this research⁽¹⁹⁾. This system was designed to select experts for validation studies in nursing and can be adapted for experts in other areas. Based on the model's system (Chart 1), experts scoring at least 5 points are eligible.

The inclusion criteria for students were being regularly enrolled in the SLH program at the originating institution and having taken the subject in voice whose content includes training in spectrographic analysis of the voice.

SAP is meant to be applied primarily by SLH pathologists in clinical voice assessment or research, and its target population includes SLH students (pathologists in training) and professionals (with a bachelor's degree in SLH Sciences). Hence, the study recruited participants who were either SLH pathologists or SLH students to ensure that each stratum of the target population was represented.

Professionals were recruited via e-mail sent to 10 SLH pathologists who were the research institution's undergraduate or postgraduate alumni or postgraduate professors, with a history of research or outreach activities in voice, specifically at the laboratory where the present research took place. The list with these SLH pathologists' names and contacts was provided by the coordination of the said laboratory.

The e-mail had research information, the informed consent form, and six questions related to the Fehring score criteria: "Do you have a master's degree with a dissertation in voice?"

Chart 1. Adaptation of the Fehring score system to select experts for validation studies

FEHRING CRITERIA	SCORE	ADAPTED CRITERIA	ADAPTED SCORE
Master of nursing.	4	Bachelor's degree in speech-language-hearing sciences (mandatory criterion)	0
Master of nursing – dissertation with relevant content in the clinical area.	1	Master's degree with a dissertation on voice	2
Research – published articles on diagnostics.	2	Research – published articles on voice	2
Article on diagnostics published in a reference journal.	2	Article on voice published in a journal ≥ B1	2
Doctor of diagnostics.	2	Doctoral degree with thesis on voice	2
At least 1 year of clinical experience in internal medicine nursing.	1	At least 2 years of clinical practice in voice and experience in narrowband spectrographic analysis	3
Certified in internal medicine with proven clinical experience.	2	Specialization/specialist in voice	3
Maximum score	14	Maximum score	14

Subtitle: ≥ = greater than or equal to; B1 = Qualis evaluation

“Have you ever carried out and published research in voice?”, “Have you published any article in voice in a journal rated B1 or above?”, “Do you have a doctoral thesis in voice?”, “Do you have at least two years of clinical practice in voice, with experience in narrowband spectrographic analysis?”, and “Do you have specialization in voice and/or are you a voice specialist?”. All 10 SLH pathologists (seven females and three males) agreed to participate and signed the informed consent form.

SLH students were likewise recruited via e-mail. The laboratory coordination provided a list of undergraduate students who were carrying out research or outreach programs in voice. The first 10 students on the list who met the eligibility criteria (having already taken the subject in voice with mandatory training in spectrographic analysis) were selected (nine females and one male). Then, e-mails were sent to these 10 students with research information, the ICF, and two questions to confirm the eligibility criteria: “Do you participate in research or outreach programs in voice?”; “Have you already taken the subject with training in spectrographic voice analysis?”. All 10 students agreed to participate and signed the informed consent form.

All professionals and students who confirmed their availability to participate and signed the informed consent form received another e-mail to schedule a day and time to collect data individually at the laboratory where this research was carried out.

Procedures

Data were collected in three stages: 1) preparing the material, 2) applying the questionnaire, and 3) carrying out the CI.

1) Preparing the material

One of the main objectives of the stage of response process validity is to investigate the target population’s understanding of the application and the cognitive processes involved in this application. Therefore, since SAP is to be applied by SLH pathologists during the acoustic inspection of the spectrogram, it was necessary to select and organize a set of spectrograms for the analysis of research participants.

The spectrograms were selected from the database of the laboratory where the research was carried out. The database contains information on all subjects who underwent clinical voice assessment at the said laboratory between 2012 and 2019, totaling 1,800 subjects treated at the outpatient clinic in that period, with information on vocal complaints, results of endoscopy, and auditory-perceptual evaluation of voice, an all the subjects’ vocal samples.

The laboratory also has an image bank of all narrowband spectrograms during sustained vowel production. This corresponding information was collected during the first clinical voice assessment session, before beginning voice therapy.

The spectrograms used in this research are from patients of both sexes, with and without vocal deviation, generated with Fonoview software, version 4.5, Dell all-in-one desktop, unidirectional cardioid Sennheiser microphone, model E-835, placed on a pedestal and attached to a Behringer preamplifier, model U-Phoria UMC 204. The voices were collected in an acoustically treated recording booth, with noise below 50 dB SPL, 44000 Hz sampling rate, 40 ms windows, 2.5 ms update

time, 60 dB dynamic amplitude range, 7500 Hz frequency limit, and 3-second minimum time interval. The sustained vowel [ϵ] was used as a sample of pitch and loudness in the usual speech pattern, self-selected by the patients.

Three voice-specialist SLH pathologists analyzed in consensus, using SAP, all these spectrograms in previous research⁽¹¹⁾, and their responses were transcribed in the said database. This description of the presence/absence of SAP items in each spectrogram was used as a reference to evaluate the correspondence between the intention of the SAP developers and the interpretation of the respondents participating in the present research.

Thus, the researchers selected 10 patients’ spectrograms from the database and image bank. The main criterion for choosing spectrograms was the frequency of occurrence of each of the 25 SAP items among the 10 spectrograms. Each item should have the probability of being indicated at least three times in the set of 10 spectrograms and not occurring in at least three of the 10 spectrograms. This criterion was important to ensure the occurrence and non-occurrence of the 25 items. Hence, the researchers consulted the description of the spectrograms, previously carried out and recorded in the database, and performed a new acoustic inspection of each spectrogram to confirm the presence or absence of each item.

Given that SAP has 25 items and that it was necessary to present, for each item, a spectrographic image with the absence and another one with the presence of each item, the 10 spectrograms were organized into 50 screens for later presentation. The set of 10 spectrograms was used considering that more than one item was likely to occur in each spectrogram. However, to minimize the learning effect, they were randomly allocated in the presentation sequence.

Then, the response process validity stage was developed in two phases: Phase 1 – preparation and application of the questionnaire to understand the SAP application; Phase 2 – CI.

Phase 1 – Developing and applying the questionnaire on SAP application comprehension

Since SAP is to be used by professionals during clinical voice assessment, item identification involves two steps: understanding what the item is and the ability to have a visual representation of the item to be observed in the spectrogram. Thus, a questionnaire was created in Google Forms to identify the intention of the item, with the 10 previously selected spectrograms and the 25 SAP items. On each page of the questionnaire, there was an image of a spectrographic trace, followed by the question, “Do you think item x is present in this spectrogram?”, which participants answered with either “yes” or “no”. If they had any difficulty identifying the item, they should select the option that read, “Apart from ‘YES’ or ‘NO’, select this option if you had any difficulty identifying whether this item was present in the spectrogram.” Thus, 50 screens were presented on the form, corresponding to the absence or presence of each of the 25 SAP items.

The answers to the first part of the question (“Do you think item x is present in this spectrogram?”) were used to verify the participants’ performance in identifying each SAP item in the spectrographic trace. These responses were compared with those of the reference judge, as previously described, to calculate the accuracy rate and sensitivity and specificity values. The second part of the question (“Apart from ‘YES’ or ‘NO’, select this option if you had any difficulty identifying whether this item

was present in the spectrogram”) was used to investigate the difficulty in using SAP items.

If the participant gave a negative answer to the first part of the question (“Do you think item x is present in this spectrogram?”), they did not proceed to the second part, related to usage difficulties. Therefore, the second part of the question (usage difficulties) may have fewer respondents than the total number of participants.

All participants carried out this phase individually at the laboratory where the research was conducted, taking 60 minutes on average. Upon finishing the spectrogram analysis, participants were invited to the CI in the next phase.

Phase 2 – Cognitive interview

CI is one of the main strategies in the development of an instrument to specifically evaluate content and response process validity⁽²⁰⁾. Its main objective is to identify items in which the instrument’s proponent’s intention is misaligned with the respondent’s interpretation, thus changing such items based on the participants’ responses.

CI involves identifying item intent, data collection, and the analysis of the respondent’s interpretation in comparison with the intended meaning⁽²¹⁾. Item intent refers to the goal the item was designed to achieve. Before CI, the researcher must have a description of the intention of each item in the instrument – to which the study used the original description of the items published in the content validation stage⁽¹¹⁾ and the reference judges’ analyses obtained from the database.

Data collection in CI may involve two procedures^(12,21): thinking aloud (asking respondents to describe their thinking as they read or respond to the item) and verbal probing (spontaneous or structured questions asked by the investigator immediately upon the participant’s response to the item). This research used verbal probing, as provides access to the four sources of cognitive biases that must be investigated in instrument validation: understanding, memory, judgment, and response.

Thus, the researcher asked the following question in CI regarding each SAP item: “How do you understand the command described in item x?”. The interviewer transcribed all participants’ responses, hoping to verify with them their understanding and usage difficulties regarding each item. Participants were also invited to make suggestions for each item, which were likewise transcribed for later analysis.

There is no consensus in the available literature concerning data analysis in CI⁽²¹⁾. It is recommended to check for consensus in the participants’ responses, and whether they identified the intention of the item proposed by the instrument developer. Therefore, in this research, participants’ responses to the question, “How do you understand the command described in item x?” (regarding each SAP item) were transcribed in a spreadsheet. Three researchers trained in SAP and blind to the research objective received the transcribed responses. Independently, they were asked to evaluate the participants’ responses to each item, answering with “yes” or “no” to the question, “Does the description correspond to the objective to be achieved by the item as proposed in SAP?”. These three evaluators had previous experience with SAP and access to a table describing the constructs and objectives of each SAP item and domain.

The researchers analyzed the three evaluators’ responses. All items in which at least two of them answered “yes” to the

question, “Does the description correspond to the objective to be achieved by the item as proposed in SAP?” were considered to have correct intent identification. Items that did not obtain mostly positive responses were identified for subsequent reformulation or exclusion from SAP⁽²⁰⁾.

Data analysis

The proportions between the correspondence (correctness) and non-correspondence (error) of participants’ and reference judges’ responses were analyzed to evaluate the judges’ responses regarding item identification, according to a previous consultation with the database. To this end, the chi-square test and accuracy, sensitivity, and specificity value tests were performed. Accuracy values were classified as excellent (greater than 90%), good (between 80% and 90%), acceptable (between 70% and 80%), poor (between 60% and 70%) and unacceptable performance (less than 60%)⁽²²⁾.

The chi-square test was also used to compare the frequency of difficulty in evaluating the item in the spectrograms, according to item usage responses (whether participants had difficulty identifying the absence/presence of the item in the spectrogram). Statistical analyses were performed in SPSS software, and the significance level was set at $p < 0.05$.

Lastly, the participants’ responses in the CI were analyzed, qualitatively identifying the items whose intention in SAP was correctly identified.

This research used the following criteria to reformulate or exclude items, according to the purpose of analysis results in the response process validity stage:

- Items identified by participants in the spectrograms with an accuracy rate greater than 70%.
- Items with a greater proportion of responses indicating that participants had difficulty using it, in comparison with the absence of such difficulty.
- Items which participants, during CI, suggested excluding or changing.

All items that failed at least two of these three criteria were reformulated or excluded from SAP. It was also decided to exclude the items that were so suggested by most participants during CI.

RESULTS

Phase 1 – Developing and applying the questionnaire on SAP application comprehension:

Most SAP items had a difference in the proportion of the participants’ correct and incorrect responses on item identification in relation to the reference judge’s assessment. Accuracy checks the rate of item identification in the spectrographic trace. Four items had excellent accuracy (5, 6, 7, and 10), 10 items had good accuracy (1, 8, 11, 13, 15, 16, 17, 21, 22, 23, and 24), three items had acceptable accuracy (9, 12, and 25), three items had poor accuracy (2, 4, and 18), and four items did not have acceptable accuracy (3, 14, 19, and 20) (Table 1).

Also, there was a statistically significant difference between responses of presence versus absence of usage difficulty (identifying the item in the spectrogram). Most participants did not have difficulty identifying any SAP item (Table 2).

Phase 2 – CI:

The intention of only six items (2, 5, 14, 15, 19, and 20) was not correctly identified in CI, as verified in the qualitative analysis. Moreover, participants suggested excluding five items (3, 4, 8, 17, and 25). The results in this stage of SAP validation and the decisions made regarding instrument items are summarized in Chart 2.

Table 1. Comparison of frequency distribution and accuracy rate regarding the identification of each item of the Spectrographic Voice Analysis Protocol in the spectrograms presented

SAP ITEM	SPECTROGRAM							
	PRESENT		ABSENT		p-value	Accu	Sens	Spec
	n	%	n	%				
Item 1 - Presence of noise at the beginning of the emission					<0.001*	90.0%	86.3%	94.4%
Present	19	95%	3	15%				
Absent	1	5%	17	85%				
Item 2 - Presence of dark well-defined vertical striations at the beginning of the emission					0.491	62.5%	64.7%	60.8%
Present	11	55%	6	30%				
Absent	9	45%	14	70%				
Item 3 - High number of irregular harmonics at the beginning of the emission					0.011*	50.0%	50.0%	50.0%
Present	14	70%	14	70%				
Absent	6	30%	6	30%				
Item 4 - No irregularity or observable noise at the beginning of the trace					0.002*	60.0%	70.0%	56.6%
Present	7	35%	3	15%				
Absent	13	65%	17	85%				
Item 5 - Presence of irregularities in the trace					<0.001*	94.7%	95.0%	94.0%
Present	19	95%	1	15%				
Absent	1	5%	17	85%				
Item 6 - Gradual loss of trace definition/energy					0.027*	100%	100%	100%
Present	20	100%	7	35%				
Absent	0	0	13	65%				
Item 7 - Presence of abrupt interruptions in the trace					<0.001*	92.5%	90.4%	94.3%
Present	19	95%	2	20%				
Absent	1	5%	18	90%				
Item 8 - Presence of regular spectrographic trace					<0.001*	87.5%	85.7%	89.4%
Present	18	90%	3	15%				
Absent	2	10%	17	85%				
Item 9 - Presence of harmonics with little brightness					0.039*	80.0%	83.3%	77.2%
Present	15	75%	3	15%				
Absent	5	25%	17	85%				
Item 10 - Decreased energy and number of harmonics above 4000 Hz					<0.001*	100%	100%	100%
Present	20	100%	2	10%				
Absent	0	0	18	90%				
Item 11 - Presence of harmonics above 4000 Hz					0.039*	90.0%	90.0%	90.0%
Present	18	90%	2	10%				
Absent	2	10%	18	90%				

* Significant values ($p < 0.05$) – Chi-square test

Subtitle: SAP = Spectrographic Voice Analysis Protocol; Accu = accuracy; Sens = sensitivity; Spec = specificity; n = absolute frequency in the item checked by the judges; % = percentage; < = less than; Hz = hertz; F_0 = Fundamental Frequency

Table 1. Continued...

SAP ITEM	SPECTROGRAM							
	PRESENT		ABSENT		p-value	Accu	Sens	Spec
	n	%	n	%				
Item 12 - Decreased energy or number of harmonics up to 4000 Hz					<0.001*	72.5%	68.0%	80.0%
Present	17	85%	8	40%				
Absent	3	15%	12	60%				
Item 13 - Increased energy between 1000 and 3000 Hz					<0.001*	85.0%	79.1%	93.7%
Present	19	95%	5	25%				
Absent	1	5%	15	75%				
Item 14 - Predominance of F0 trace rather than that of other harmonics					0.002*	55.0%	53.3%	60.0%
Present	16	80%	14	70%				
Absent	4	20%	6	30%				
Item 15 - Decreased energy level in the whole frequency band throughout the trace					0.012*	90.0%	90.0%	90.0%
Present	18	90%	2	10%				
Absent	2	10%	18	90%				
Item 16 - Increased energy level in the whole frequency band throughout the trace					<0.001*	90.0%	90.0%	90.0%
Present	18	90%	1	5%				
Absent	2	10%	19	95%				
Item 17 - Series rich in regular harmonics up to 4000 Hz					0.058	85.0%	100%	76.9%
Present	14	70%	0	0				
Absent	6	30%	20	100%				
Item 18 - Presence of irregular horizontal striations between harmonics					0.012*	66.6%	70.0%	62.5%
Present	14	70%	10	50%				
Absent	6	30%	10	50%				
Item 19 - Presence of undefined or sketched harmonics					0.004*	57.5%	55.1%	63.63%
Present	16	80%	13	65%				
Absent	4	20%	7	35%				
Item 20 - Predominant presence of low-amplitude harmonics					<0.001*	57.1%	55.0%	59.1%
Present	11	55%	7	35%				
Absent	9	45%	13	65%				
Item 21 - Presence of irregular harmonics					<0.001*	86.0%	80.9%	89.6%
Present	17	85%	4	20%				
Absent	3	15%	26	80%				
Item 22 - Presence of noise between harmonics below 4000 Hz					0.042*	80.0%	77.2%	83.3%
Present	17	85%	5	25%				
Absent	3	15%	15	75%				
Item 23 - Presence of additional diffuse noise above 4000 Hz					<0.001*	87.5%	85.7%	89.4%
Present	18	90%	3	15%				
Absent	2	10%	17	85%				
Item 24 - Harmonics replaced with noise in the spectrographic trace					<0.001*	85.0%	79.1%	93.7%
Present	19	95%	5	25%				
Absent	1	5%	15	75%				
Item 25 - Spectrographic trace with discrete or no noise between harmonics below 4000 Hz					0.039*	77.5%	73.9%	82.3%
Present	17	85%	6	30%				
Absent	3	15%	14	70%				

* Significant values ($p < 0.05$) – Chi-square test

Subtitle: SAP = Spectrographic Voice Analysis Protocol; Accu = accuracy; Sens = sensitivity; Spec = specificity; n = absolute frequency in the item checked by the judges; % = percentage; < = less than; Hz = hertz; F₀ = Fundamental Frequency

Table 2. Frequency distribution regarding difficulties using each item of the Spectrographic Voice Analysis Protocol in the spectrograms presented

SAP ITEM	USAGE DIFFICULTIES				p-value
	PRESENT		ABSENT		
	n	%	n	%	
Item 1 - Presence of noise at the beginning of the emission	1	2.63%	18	97.37%	<0.001*
Item 2 - Presence of dark well-defined vertical striations at the beginning of the emission	5	13.15%	14	86.85%	0.002*
Item 3 - High number of irregular harmonics at the beginning of the emission	4	10.52%	15	89.48%	<0.001*
Item 4 - No irregularity or observable noise at the beginning of the trace	5	13.15%	14	86.85%	<0.001*
Item 5 - Presence of irregularities in the trace	0	0.00%	19	100%	<0.001*
Item 6 - Gradual loss of trace definition/energy	1	2.63%	18	97.37%	0.002*
Item 7 - Presence of abrupt interruptions in the trace	0	0.00%	19	100%	<0.001*
Item 8 - Presence of regular spectrographic trace	3	7.89%	16	92.11%	<0.001*
Item 9 - Presence of harmonics with little brightness	4	10.52%	15	89.48%	0.025*
Item 10 - Decreased energy and number of harmonics above 4000 Hz	2	5.26%	17	94.74%	<0.001*
Item 11 - Presence of harmonics above 4000 Hz	1	2.63%	18	97.37%	<0.001*
Item 12 - Decreased energy or number of harmonics up to 4000 Hz	2	5.26%	17	94.74%	<0.001*
Item 13 - Increased energy between 1000 and 3000 Hz	1	2.63%	18	97.37%	0.007*
Item 14 - Predominance of F0 trace rather than that of other harmonics	3	7.89%	16	92.11%	0.002*
Item 15 - Decreased energy level in the whole frequency band throughout the trace	3	7.89%	16	92.11%	<0.001*
Item 16 - Increased energy level in the whole frequency band throughout the trace	1	2.63%	18	97.37%	<0.001*
Item 17 - Series rich in regular harmonics up to 4000 Hz	1	2.63%	18	97.37%	0.007*
Item 18 - Presence of irregular horizontal striations between harmonics	6	15.78%	13	84.22%	<0.001*
Item 19 - Presence of undefined or sketched harmonics	0	0.00%	19	100%	<0.001*
Item 20 - Predominant presence of low-amplitude harmonics	7	18.41%	12	81.59%	0.002*
Item 21 - Presence of irregular harmonics	2	5.26%	17	94.74%	0.025*
Item 22 - Presence of noise between harmonics below 4000 Hz	1	2.63%	18	97.37%	<0.001*
Item 23 - Presence of additional diffuse noise above 4000 Hz	2	5.26%	17	94.74%	<0.001*
Item 24 - Harmonics replaced with noise in the spectrographic trace	1	2.63%	18	97.37%	<0.001*
Item 25 - Spectrographic trace with discrete or no noise between harmonics below 4000 Hz	5	13.15%	14	86.85%	<0.001*

* Significant values (p < 0.05) – Chi-square test

Subtitle: SAP = Spectrographic Voice Analysis Protocol; n = absolute frequency in the item checked by the judges; % = percentage; < = less than; Hz = hertz; F₀ = Fundamental Frequency

Chart 2. Items excluded or reformulated after the cognitive interview and application of the Spectrographic Voice Analysis Protocol

Item	Domain	Decision	Criteria
(2) Presence of dark well-defined vertical striations at the beginning of the emission	Beginning of the emission	Excluding the item	Accuracy ≤ 70% Difficulties identifying the intention of the item
(3) High number of irregular harmonics at the beginning of the emission	Beginning of the emission	Excluding the item	Accuracy ≤ 70% Suggested in the cognitive interview
(4) No irregularity or observable noise at the beginning of the trace	Beginning of the emission	Excluding the item	Accuracy ≤ 70% Suggested in the cognitive interview
(5) Presence of irregularities in the trace	Temporal aspects of the emission	Reformulating the item as follows: "Changes in the spectrographic trace configuration in the time domain"	Difficulties identifying the intention of the item
(8) Presence of regular spectrographic trace	Temporal aspects of the emission	Excluding the item	Suggested in the cognitive interview
(14) Predominance of F0 trace rather than that of other harmonics	Energy distribution in the trace	Excluding the item	Accuracy ≤ 70% Difficulties identifying the intention of the item
(15) Decreased energy level in the whole frequency band throughout the trace	Energy distribution in the trace	Maintaining the item unchanged	Difficulties identifying the intention of the item
(17) Series rich in regular harmonics up to 4000 Hz	Energy distribution in the trace	Excluding the item	Suggested in the cognitive interview
(18) Presence of irregular horizontal striations between harmonics	Description of the harmonics	Maintaining the item unchanged	Accuracy ≤ 70%
(19) Presence of undefined or sketched harmonics	Description of the harmonics	Maintaining the item unchanged	Accuracy ≤ 70% Difficulties identifying the intention of the item
(20) Predominant presence of low-amplitude harmonics	Description of the harmonics	Maintaining the item unchanged	Accuracy ≤ 70% Difficulties identifying the intention of the item
(21) Presence of irregular harmonics	Description of the harmonics	Reformulating the item as follows: "Presence of harmonics with irregular trajectory and morphology (non-rectilinear)"	Suggested in the cognitive interview
(25) Spectrographic trace with discrete or no noise between harmonics below 4000 Hz	Distribution of noise in the trace	Excluding the item	Suggested in the cognitive interview

Subtitle: ≤ = less than or equal to; F₀ = Fundamental Frequency

DISCUSSION

The process of applying a questionnaire or tool that demands the subject's response requires them to engage in four cognitive operations: understanding, memory, judgment, and response⁽²³⁾. Thus, respondents must understand the question asked, retrieve relevant information or knowledge from memory, make a judgment about the item or information retrieved, and select an answer. Each of these operations is a source of variability in tool application and, consequently, a potential source of error regarding the reliability of responses obtained with the tool. Therefore, in the process of developing an instrument, it must be checked whether the respondent's interpretation corresponds to what the instrument developer intends with each item⁽²⁴⁾. This is evaluated in the response process validation stage.

This research aimed to carry out the SAP response process validation stage, which was achieved and culminated in the reformulation of the SAP. Therefore, the SAP version presented at the end of this study has response process validity, indicating that users (SLH pathologists and SLH students) understood the SAP application and analysis.

The initial version (before this validation stage) had 25 items, structured into five domains. Based on the research results, two items were reformulated (5 and 21) and seven were excluded (2, 3, 4, 8, 14, 17, and 25), resulting in a new SAP version whose 18 items have response process validity.

SAP is meant to be applied by the clinician, not the patient, based on the acoustic inspection of the spectrogram. The objective is to standardize spectrogram descriptions for characterizing dysphonic and non-dysphonic voices and improve communication between clinicians and researchers using spectrography. Therefore, the professionals' interpretation is essential for all inferences based on SAP application results, whether in clinical use, evaluating and monitoring dysphonic patients, or in research.

This reinforces the relevance of the validation stage approached in this research, as it identifies the professionals' difficulties regarding SAP items in their understanding, memory, judgment, and response⁽²³⁾. Such identification also leads to excluding or changing items for a new version of the tool.

Items 2, 3, 4, 14, 18, 19, and 20 had an accuracy rate below or equal to 70% in their identification during acoustic inspection of the spectrogram. This may indicate the participants' difficulty in understanding the item and searching their memory for a reference to identify it in the spectrographic trace, which leads to further errors. Of these items, participants also had difficulty identifying the intention of items 2, 14, 19, and 20. This difficulty may reflect the misalignment between the SAP developers' intention and the research participant's interpretation.

Besides their accuracy rate below or equal to 70%, participants suggested excluding items 3 and 4 from SAP due to the difficulty in characterizing them in the spectrogram. Items in a tool that are inappropriately identified (when the tool construct includes expected responses or reference values) or misaligned with the developer's original intention need to be excluded or reformulated⁽²¹⁾.

Item 18 had an accuracy rate below or equal to 70%, although participants did not have difficulty identifying its intention, nor

did they suggest its reformulation or exclusion. Therefore, it was maintained without changes in the final SAP version. The presence of irregular horizontal striations between harmonics is classically described^(5,10) as one of the main components for identifying dysphonic voices in the spectrographic trace. Nomenclatures commonly describe this item as "horizontal streaks between harmonics"^(10,11), "harmonic bifurcation"⁽⁵⁾, or "sub-harmonics"⁽⁵⁾. Thus, the low accuracy (66.6%) found in this study regarding the identification of the presence/absence of this item in the spectrographic trace can be justified by the different possible names to describe such a characteristic in a spectrogram.

Despite its low accuracy, item 18 met the other two eligibility criteria for maintenance in the SAP version at the end of this research. Naturally, the process of training clinicians and researchers to apply SAP should use sample spectrograms with the description of this item to facilitate its correct identification.

The main objective of response process validation studies during the development of a test or questionnaire is to understand whether the respondents' cognitive processes, activated during the application of the test or questionnaire, are aligned with the instrument developers' intention⁽²⁵⁾. Hence, at the end of this validation stage, the researchers who developed the test or questionnaire are responsible for establishing the criteria for maintaining items without changing, reformulating, or deleting them^(14,21,26).

Participants had difficulty identifying the intention in items 5 and 15. Since these items did not meet the other exclusion criteria, they were maintained in the final version. On the other hand, as the researchers who developed the tool were responsible for evaluating whether to maintain, exclude, or change an item in the final SAP version, they decided to reformulate item 5. Thus, it was described as "Changes in the spectrographic trace configuration in the time domain". It was understood that the original description of item 5 ("Presence of irregularity in the trace") was non-specific, as it did not refer to either the morphology of the harmonics or their trajectory over time. In turn, item 15 was maintained without reformulation.

During the CI, participants suggested the reformulation of item 21. Therefore, in the final SAP version, it was reformulated to "Presence of harmonics with irregular trajectory and morphology (non-rectilinear)". Participants commented that there could be confusion between items 5 and 21. Hence, both were changed in the final SAP version.

Participants also suggested excluding items 8, 17, and 25. Therefore, based on this suggestion and the SAP developers' evaluation of their clinical relevance, these items were excluded from the final SAP version. At the end of this research, a new SAP version was developed, which is shown in Figure 1.

In general, the results in this study support the response process validity of SAP. A limitation of the present study was that all participants had not completed precisely the same training on SAP before participating. Both SLH pathologists and SLH students had prior training in spectrographic analysis; the researchers had no control over the conditions of such training and the participants' (either students or pathologists) actual proficiency in acoustic inspection of the spectrogram.

In turn, SAP terminology is the same widely used in the area, which may have led to good results in the participants' understanding of the items. This limitation does not invalidate the results of the present study, but it can be addressed in future studies.

CONCLUSION

Expert and non-expert participants had an accuracy rate greater than 70% in identifying most SAP items in the spectrogram. Only seven items had an accuracy below or equal to 70% in the participants' assessment. Regarding usage difficulties, most participants had no difficulty identifying all SAP items. In the CI, the intention of only six items was not correctly identified, as verified in the qualitative analysis. Furthermore, participants suggested excluding five items. Thus, the final SAP version, after this stage, was reduced from 25 to 18 items.

REFERENCES

- Nemr K, Amar A, Abrahão M, Leite GCDA, Köhle J, Santos ADO, et al. Análise comparativa entre avaliação fonoaudiológica perceptivo-auditiva, análise acústica e laringoscópias indiretas para avaliação vocal em população com queixa vocal. *Rev Bras Otorrinolaringol.* 2005;71(1):13-7. <http://dx.doi.org/10.1590/S0034-72992005000100003>.
- Barsties B, De Bodt M. Assessment of voice quality: current state-of-the-art. *Auris Nasus Larynx.* 2015;42(3):183-8. <http://dx.doi.org/10.1016/j.anl.2014.11.001>. PMID:25440411.
- Lopes L, Cavalcante D. Intensidade do desvio vocal: integração de dados perceptivo-auditivos e acústicos em pacientes disfônicos. *CoDAS.* 2014;26:382-8. <http://dx.doi.org/10.1590/2317-1782/20142013033>. PMID:25388071.
- Eadie TL, Doyle PC. Classification of dysphonic voice: acoustic and auditory-perceptual measures. *J Voice.* 2005;19(1):1-14. <http://dx.doi.org/10.1016/j.jvoice.2004.02.002>. PMID:15766846.
- Titze IR. Workshop on Acoustic Voice Analysis: Summary Statement [Internet]. 1995 [citado em 2022 Fev 7]. Disponível em: https://scholar.google.com/scholar?cluster=16280338619419163408&hl=en&as_sdt=2005&sciodt=0,5
- Brockmann-Bausier M, Drinnan MJ. Routine acoustic voice analysis: time to think again? *Curr Opin Otolaryngol Head Neck Surg.* 2011;19(3):165-70. <http://dx.doi.org/10.1097/MOO.0b013e32834575fe>. PMID:21483265.
- Christmann MK, Brancalioni AR, Freitas CRD, Vargas DZ, Keske-Soares M, Mezzomo CL, et al. Uso do programa MDVP em diferentes contextos: revisão de literatura. *Rev CEFAC.* 2015;17(4):1341-9. <http://dx.doi.org/10.1590/1982-021620151742914>.
- Lopes LW, Alves GÂDS, Melo MLD. Content evidence of a spectrographic analysis protocol. *Rev CEFAC.* 2017;19(4):510-28. <http://dx.doi.org/10.1590/1982-021620171942917>.
- Bastilha GR, Pagliarini KC, Moraes DAO, Cielo CA. Spectrographic Vocal Assessment Protocol (SVAP): Reliability and Criterion Validity. *J Voice.* 2021 Nov 1;35(6):931.e1-14. <http://dx.doi.org/10.1016/j.jvoice.2020.02.017>. PMID:32209278.
- Yanagihara N. Significance of harmonic changes and noise components in hoarseness. *J Speech Hear Res.* 1967;10(3):531-41. <http://dx.doi.org/10.1044/jshr.1003.531>. PMID:6081935.
- Lopes LW, Alves GÂDS, Melo MLD. Content evidence of a spectrographic analysis protocol. *Rev CEFAC.* 2017;19(4):510-28. <http://dx.doi.org/10.1590/1982-021620171942917>.
- Pernambuco L, Espelt A, Magalhães H, Lima KC. Recommendations for elaboration, transcultural adaptation and validation process of tests in Speech, Hearing and Language Pathology. *CoDAS.* 2017 Jun 8;29(3):e20160217. PMID:28614460.
- Gonçalves MIR, Pontes PADL, Vieira VP, Pontes AADL, Curcio D, Biase NGD. Função de transferência das vogais orais do Português brasileiro: análise acústica comparativa. *Rev Bras Otorrinolaringol (Engl Ed).* 2009;75:680-4.
- García JP, Baena IB. Validity evidence based on response processes. *Psicothema.* 2014;26(1):136-44. PMID:24444741.
- Plake BS, Wise LL. What Is the Role and Importance of the Revised AERA, APA, NCME Standards for Educational and Psychological Testing? *Educ Meas.* 2014 Dez 1;33(4):4-12. <http://dx.doi.org/10.1111/emip.12045>.
- Boateng GO, Neilands TB, Frongillo EA, Melgar-Quinonez HR, Young SL. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Front Public Health.* 2018 Jun 11;6:149. <http://dx.doi.org/10.3389/fpubh.2018.00149>. PMID:29942800.
- Beatty PC, Willis GB. Research synthesis: the practice of cognitive interviewing. *Public Opin Q.* 2007;71(2):287-311. <http://dx.doi.org/10.1093/poq/nfm006>.
- Blair J, Conrad FG. Sample size for cognitive interview pretesting. *Public Opin Q.* 2011;75(4):636-58. <http://dx.doi.org/10.1093/poq/nfr035>.
- Fehring RJ. The fehring model. In: *Classification of Nursing Diagnoses.* Philadelphia: JB Lippincott; 1994. p. 55-62.
- Castillo-Díaz M, Padilla JL. How cognitive interviewing can provide validity evidence of the response processes to scale items. *Soc Indic Res.* 2013 Dez;114(3):963-75. <http://dx.doi.org/10.1007/s11205-012-0184-8>.
- Peterson CH, Gischlar KL, Peterson NA. Item construction using reflective, formative, or rasch measurement models: implications for group work. *J Spec Group Work.* 2017 Jan 2;42(1):17-32. <http://dx.doi.org/10.1080/01933922.2016.1264523>.
- Hosmer DW, Lemeshow S. *Applied logistic regression.* New York: Wiley; 2000. <http://dx.doi.org/10.1002/0471722146>.
- Tourangeau R. Cognitive science and survey methods: a cognitive perspective. In: Tourangeau R, editor. *Cognitive aspects of survey design: building a bridge between disciplines.* Washington: National Academy Press; 1984. p. 73-100.
- Ryan K, Gannon-Slater N, Culbertson MJ. Improving survey methods with cognitive interviews in small- and medium-scale evaluations. *Am J Eval.* 2012;33(3):414-30. <http://dx.doi.org/10.1177/1098214012441499>.
- Padilla JL, Benítez I. Validity evidence based on response processes. *Psicothema.* 2014;26(1):136-44. PMID:24444741.
- Hawkins M, Elsworth GR, Hoban E, Osborne RH. Questionnaire validation practice within a theoretical framework: a systematic descriptive literature review of health literacy assessments. *BMJ Open.* 2020 Jun 1;10(6):e035974. <http://dx.doi.org/10.1136/bmjopen-2019-035974>. PMID:32487577.