RevistaCEFAC
Current Evidence on Feeding, Audiology and Communication

eISSN 1982-0216

DOI: 10.1590/1982-0216/20252742024 | Rev. CEFAC. 2025;27(4):e2024

Original articles

Auditory-perceptual evaluation of the overall severity of vocal deviation by comparing the visual analog and numerical scales

Priscila Campos Martins dos Santos[1]
Ana Cristina Côrtes Gama[1]

[1] Universidade Federal de Minas Gerais - UFMG, Faculdade de Medicina, Departamento de Fonoaudiologia, Belo Horizonte, Minas Gerais, Brasil.

A study conducted at the Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil.

Financial support: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES): Finance Code 001 (PDPG do Programa de Pós-Graduação em Ciências Fonoaudiológicas) and process number 88887.615544/2021-00

Conflict of interests: Ana Cristina Côrtes Gama declares she is an editorial board member of Revista CEFAC but was not involved in the peer review or editorial decision-making process for this article

Corresponding author:
Priscila Campos Martins dos Santos
Avenida Professor Alfredo Balena, 190/ Sala 249 - Santa Efigênia
CEP: 30130-100 - Belo Horizonte, MG, Brasil
E-mail: priscila.fonoaudiologia@gmail.com

Received on February 27, 2024
Received in a revised form on August 27, 2024
Accepted on February 27, 2025

Chief Editor: Renata Furlan

ABSTRACT

Purpose: to analyze the cutoff points, sensitivity, and specificity of the correspondence between the visual analog and numerical scales in assessing the overall severity of vocal deviation with linked speech tasks.

Methods: 105 voices were selected for the linked speech task. Six speech-language-hearing pathologists evaluated the samples regarding the overall severity of deviation, using the visual analog and numerical scales, with a 2-day interval between them. The study defined cutoff values based on sensitivity and specificity, used the intraclass correlation coefficient, kappa coefficient, Spearman's coefficient, analysis of variance, and analysis of sensitivity and specificity, and set the significance level at 5%.

Results: the cutoff values were defined as absent deviation – up to 27 mm, mild – from 27.1 to 50.5 mm, moderate – from 50.6 to 68.5 mm, and intense deviation – 68.6 to 100 mm. The following values were found regarding performance, sensitivity, and specificity, respectively: absent deviation - 0.74, 0.94, and 0.86; mild - 0.57, 0.92, and 0.85; moderate - 0.68, 0.93, and 0.89; intense - 0.95, 0.95, and 0.95.

Conclusion: the numerical scale adequately identified the cutoff points of the overall severity of vocal deviation on the visual analog scale, during the linked speech task.

Keywords: Voice; Voice Quality; Dysphonia; Auditory Perception; Voice Training

INTRODUCTION

Auditory-perceptual evaluation is the main speech-language-hearing assessment in voice clinical practice[1]. This is a subjective assessment, as it depends on the evaluator's auditory impression of the patient's voice, subject to extrinsic and intrinsic influences[1,2]. The sample presentation forms, the evaluation scales, and the speaking tasks stand out among the extrinsic influences[2]. Several aspects can interfere with the evaluator's intrinsic performance: the type and length of experience and professional training, auditory training, attention to the evaluation, and the mother tongue[2]. Despite its subjectivity, some strategies increase the auditory-perceptual evaluation robustness, such as using scales in the assessment[3].

The listener's references influence the perception of a voice as healthy or changed, increasing the variability of this assessment[1]. The use of validated and widely used protocols, such as the CAPE-V visual analog scale (VAS) and the GRBAS numerical scale (NS), in voice clinical practice and research reduces this variability[1]. Currently, CAPE-V and GRBAS are the most widely used scales in voice clinical practice and research[1].

CAPE-V[4] consists of a VAS (i.e., a 100-mm straight line) in which listeners mark the point corresponding to the intensity of deviation of a given parameter, with 0 mm representing no deviation and 100 mm representing the maximum level of deviation. GRBAS[5] is a 4-point NS, through which the voice is classified as 0 – no deviation, 1 – slight deviation, 2 – moderate deviation, or 3 – intense deviation.

Although both scales are widely used in voice clinical practice, their different measurement forms hinder a direct comparison between them. The VAS is more sensitive in capturing the variability of vocal quality[1]. However, to define cutoff points in a VAS, understanding its representativeness for a healthy or changed voice, and the degree of vocal deviation for a given parameter, studies must analyze its correspondence with an NS[1]. Studies that analyze the correspondence between these scales[6-12] used generic VAS and NS since they include the CAPE-V and GRBAS without being limited to them[11].

The GRBAS and CAPE-V also differ regarding the speaking task requested of the speaker. GRBAS does not present specific speaking tasks[1,13]. However, clinical practice commonly uses sustained vowel tasks, linked speech, and spontaneous speech to assess vocal quality using the GRBAS. CAPE-V uses sustained vowels, sentences, and spontaneous conversations as speech tasks. Much of the literature refers to sustained vowels and linked speech as speech tasks[1,14-15]. Sustained vowels allow for an analysis of vocal quality at the glottal level, since their emission is less affected by the vocal tract. Linked speech, on the other hand, carries the effects of articulation, resonance, and prosody – i.e., its emission is affected by the interference of the vocal tract. Therefore, linked speech is closer to natural emission for the listener and provides information on habitual voice patterns used in everyday communication situations[1]. This perception of how speech tasks impact voice analysis indicates that both should be used to assess vocal quality[1,14-17]. The comparison between different speech tasks – sustained vowels, linked speech, and spontaneous speech – is fundamental for the auditory-perceptual evaluation of voice, providing distinct information about vocal quality and the functional aspects of voice production. Although more studies in the literature have analyzed the correspondence between VAS and NS based on linked speech[6-10] than sustained vowel tasks[11-12], important differences are noted in the cutoffs, requiring continued studies of the scales based on linked speech tasks.

The scientific community has advanced with the development of correspondence points between the two scales, VAS and NS[6-12]. Early studies defined a cutoff in the VAS for normal and changed voices[6,7] with linked speech tasks. The literature brings new studies on the levels of deviation for the overall severity of vocal deviation[8], on cutoff values for older adults[9], using samples in other languages[10], but all so far for linked speech tasks. The first correspondence between the scales was carried out in 2015 with sustained vowel tasks[11] for the overall severity of vocal deviation and, subsequently, for roughness and breathiness[12].

The literature states that factors such as speech task (e.g., sustained vowel and linked speech), severity of vocal deviation, type of parameter evaluated, and linguistic factors can influence the auditory-perceptual evaluation results[1,2,3]. Although the literature has advanced in studies on the correspondence between VAS and NS, research using different tasks and populations with different mother tongues or analyzing different parameters obtains varied findings, including the cutoffs established for VAS[6-12]. Therefore, new research is needed to bring new evidence, considering different samples, populations, tasks, and analyses to validate the findings in the literature.

Thus, this study aimed to analyze the cutoff points, sensitivity, and specificity of the correspondence between VAS and NS in evaluating the overall severity of vocal deviation with linked speech tasks.

METHODS

This cross-sectional, analytical, observational, quantitative study was approved by the Research Ethics Committee of the Universidade Federal de Minas Gerais, MG, Brazil, under evaluation report number 4.812.704 and CAEE number 42739021.2.1001.5149. All study participants read, agreed to, and signed an informed consent form.

For the initial selection of linked speech samples, two judges – speech-language-hearing pathologists with at least 10 years of experience in auditory-perceptual evaluation of voice – analyzed 381 voices from a speech-language-hearing outpatient clinic database. The database consisted of linked speech samples – days of the week: Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday; samples of men and women aged 18 years or older; healthy and changed voices of varying degrees of deviation. For the initial selection, 20% of the voices in the database were randomly repeated for subsequent intrarater agreement analysis, which was classified as excellent[18] for both the VAS (0.89 and 0.92) and NS (0.85 and 0.87). The judges listened to and individually classified the voices in the bank according to the overall severity of vocal deviation as 0 – no deviation, 1 – slight deviation, 2 – moderate deviation, and 3 – severe deviation. The voices on which both judges agreed regarding the severity of deviation were selected, totaling 105 voices, with and without vocal deviation, with different degrees of deviation; 79 vocal samples were from women and 26 from men. Also, 20% of the sample was randomly repeated for later intrarater agreement analysis, totaling 126 samples.

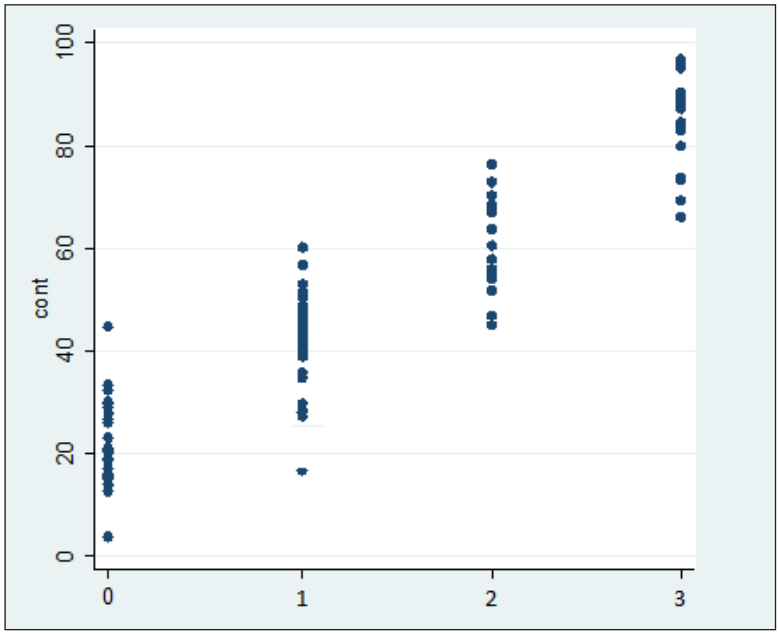
Seven judges were selected to analyze the voices using both scales. They were speech-language-hearing pathologists with over 10 years of experience in auditory-perceptual evaluation. One was excluded for not completing the voice analysis, resulting in six judges. The parameter evaluated in each voice was the overall severity of vocal deviation (i.e., the intensity of the voice deviation) using two assessment instruments (the VAS and NS). The judges received a file with the samples and another with the scales, in addition to a stereo Multilaser Vibe Headphone. Each judge evaluated the voices individually. The analysis was performed in two blocks: 1. Evaluation of the overall severity of vocal deviation using the VAS; 2. Evaluation of the overall severity of vocal deviation using the NS. There was a 2-day interval between blocks to avoid the memorization effect. The 126 voice samples were evaluated on a single day in each block, randomizing the order in which the voices were presented for each scale.

The judges were asked to listen to the sample and mark the measure equivalent to the overall severity of deviation for that voice in the VAS' 100-mm line, with 0 mm being no deviation and 100 mm being maximum deviation. As for the NS, the judges listened to the sample and classified the overall severity of vocal deviation on a 4-point scale: 0 – no deviation; 1 – slight deviation; 2 – moderate deviation; 3 – severe deviation. They could listen to the voices as many times as they considered necessary.

The intrarater and interrater agreement was analyzed with the intraclass correlation coefficient for the VAS and the weighted kappa correlation coefficient for the NS. The Shapiro-Wilk test was performed after descriptive data analysis for VAS (mean, quartiles, minimum, maximum, and standard deviation) and NS (frequency and proportions), demonstrating normality of the data, and the assumptions of homogeneity for ANOVA were adjusted by the Levene test. The correspondence between the scales was evaluated by the Spearman's coefficient, considering a p-value between 0 and 0.3 – negligible correlation; between 0.31 and 0.5 – weak correlation; between 0.51 and 0.7 – moderate correlation; between 0.71 and 0.9 – strong correlation; and greater than 0.9 – very strong correlation[20]. This test was chosen due to the ordered categorical variable. The analysis of variance (ANOVA) compared the means of the continuous variable (VAS) between the levels of the categorical variable (NS). Sensitivity and specificity were verified based on true negative (TN), true positive (TP), false negative (FN), and false positive (FP) values to define the cutoff points of correspondence between the VAS and the NS. The NV, TP, FN, and FP values were found from the NS intersection with the VAS quintiles. Sensitivity was analyzed with the formula TP/(TP+FN), and specificity with NV/(NV+FP). The following classification was considered to analyze the performance measures: excellent – above 0.75, satisfactory – between 0.4 and 0.75, poor – below 0.4[18]. Statistical analysis was performed using STATA software, version 12.0, considering a significance level of 5% – p-values were significant if less than 0.05.

RESULTS

A high correspondence[20] was found between the VAS and NS (Spearman = 0.95), with symmetrical distribution between them (Figure 1).



Caption: cont = continuous – visual analog scale.

Figure 1. Scatter plot with the categorical scale vs. the continuous scale

There was greater intrarater (ICC above 0.8 for all evaluators) and interrater (ICC – average 0.9) agreement for the VAS (intrarater values ranging from 0.884 to 0.975) than for the NS (intrarater values ranging from 0.548 to 0.751) – Table 1.

Table 1. Intrarater and interrater agreement values for the visual analog and numerical scales

	VAS	NS
Mean of the intrarater agreement	0.927	0.660
Overall interrater agreement	0.906	0.186

Captions: VAS = visual analog scale; NS = numerical scale.
The agreement was calculated with the Intraclass Correlation Coefficient for the VAS and the weighted kappa coefficient for the NS
.

The descriptive data analysis from the correspondence between the VAS and NS is found in Table 2, which presents the number of voices classified per severity of deviation, the minimum and maximum values of voice classification by the evaluators, and the mean, median, standard deviation, first and third quartiles, and the comparison between the VAS and NS means.

Table 2. Descriptive analysis of the correspondence between the visual analog and numerical scales per degree of deviation

Category	N	Min	Q1	Median	Q3	Max	Mean	Standard Deviation	p-value
0	25	3.67	15.17	19.00	23.00	26.67	18.67	5.49	< 0.001
1	38	27.83	42.50	45.00	47.67	50.50	45.15	3.60	
2	19	51.33	54.00	57.83	66.17	68.50	59.14	6.04	
3	23	69.33	73.67	84.33	90.33	96.83	84.06	9.06	

Captions: 0 = no deviation; 1 = mild deviation; 2 = moderate deviation; 3 = severe deviation. Comparison between the means of the visual analog scale and the degrees of vocal deviation with the numerical scale using the ANOVA test.

The cutoffs for the VAS in evaluating the overall severity of vocal deviation with the linked speech task were established based on the analysis of the 3rd quartile per group. These points were evaluated using sensitivity and specificity values, as described in Table 3. There was high sensitivity and specificity, and the latter had higher values than the former, except for the cutoff corresponding to degree 3.

Table 3. Analysis of cutoff points, sensitivity, and specificity of the correspondence between the visual analog and numerical scales

Degree	Cutoff points	Sensitivity	Specificity
0	27	0.74	0.94
1	50.5	0.57	0.92
2	68.5	0.68	0.93
3	100	0.95	0.95

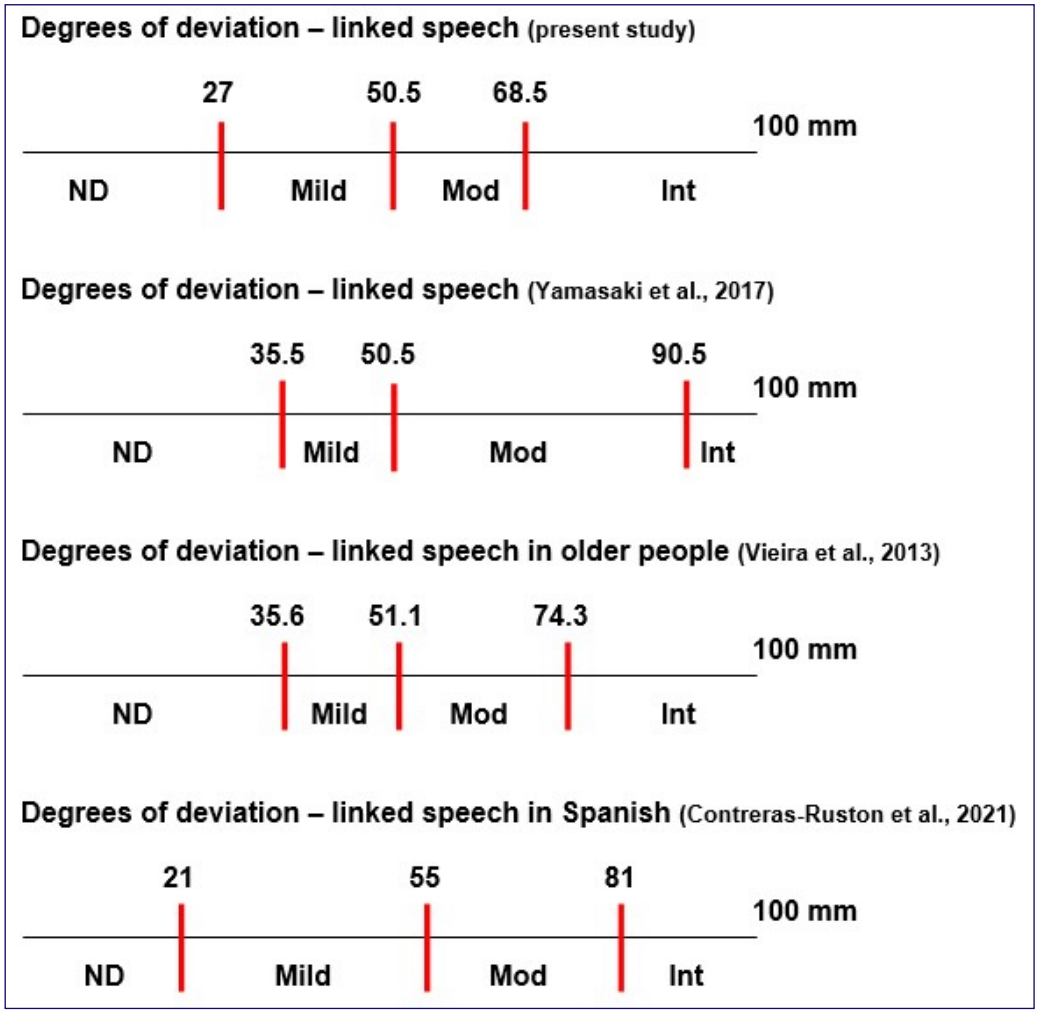
Captions: 0 = no deviation; 1 = mild deviation; 2 = moderate deviation; 3 = intense deviation.

DISCUSSION

The VAS, for being a 100-mm scale, enables greater possibilities for marking vocal deviation than the NS. However, it does not present limits for the normal variability of the voice or the different degrees of deviation[1]. The study of the correspondence between the VAS and NS makes it possible to define cutoffs, enabling descriptive analysis and voice classification as healthy or changed, and defining the degree of vocal deviation based on the VAS.

This study found a high correspondence between the VAS and NS (0.95), which corroborates the literature[4,11], and allows a comparison between their findings. It also found a higher interrater agreement for the VAS than for the NS, which can be observed in other studies[3,10]. The VAS was more sensitive to small differences in vocal deviations than the EN. Intrarater agreement was also higher for the VAS, indicating greater consistency among raters when using this scale[11]. Both intrarater and interrater agreements were higher for extreme degrees (no deviation and intense deviation) than intermediate degrees (mild and moderate deviation). This corroborates the literature, which shows greater ease for raters to analyze samples without deviation or with extreme deviation[1].

The cutoff values found in the present study for overall severity of vocal deviation with the linked speech task were close to the cutoffs of previous studies[8-10] (Figure 2).



Captions: ND = no deviation; Mod = moderate deviation; Int = intense deviation.

Figure 2. Rulers with visual analog scale graduations based on their respective cutoff values

The literature highlights that the mother tongue and age impact the auditory-perceptual evaluation[1,19]. Studies[8-10] have, like this one, evaluated the overall severity of vocal deviation with linked speech tasks, although in different populations, ages, and mother tongues – only adults aged 19 to 60 years[8], only older adults[9], and Spanish speakers[10]. Their cutoffs are very close for moderate deviations, but they present a greater difference for extreme deviations (mild and intense). This finding reinforces the impact of the mother tongue and age on the auditory-perceptual evaluation of the overall severity of vocal deviation by VAS. The present study's sample comprised a population over 18 years old, including adults and older adults. Interestingly, there is an important difference in the cutoff for intense vocal deviation between the study only with adults (90.5)[8] and the one only with older people (74.3)[9]. The cutoff for intense deviation in the present study (68.5) is closer to the one for older people[9], which can be explained by the presence of this age group in the sample. It is important to emphasize that these studies[8,9] are similar to the present one regarding other sample characteristics: their native language is Brazilian Portuguese, the sample included individuals of both sexes, with and without vocal changes in varying degrees of deviation, and their overall severity of vocal deviation was assessed with a speech task. In other words, they differ mainly regarding age, reinforcing its possible impact on the auditory-perceptual evaluation.

A correlation study between NS and VAS to assess the overall severity of vocal deviation using a sustained vowel task[11] found the following cutoffs: absent deviation – 0 to 34 mm, mild deviation – 34.1 to 51 mm, moderate deviation – 51.1 to 63.5 mm, intense deviation – 63.6 to 77.5 mm, and extreme deviation – above 77.5 mm. It found values quite close to the present

study's cutoffs with the linked speech task. According to the literature, the sustained vowel presents greater deviation than linked speech[1]. Future studies on the correspondence between NS and VAS to define cutoffs should use samples of both tasks performed by the same individuals to investigate the scales' behavior with the different speech tasks.

Studies that analyze the correspondence between VAS and NS, considering variables such as age, native language, sex, and specific audiences (teachers, singers, telemarketers, and so on), are greatly relevant for voice clinical practice. These studies should also include other tasks (e.g., sustained vowels) and parameters (e.g., roughness and breathiness). Such studies help standardize cutoffs for VAS, increasing the reliability of the auditory-perceptual evaluation using this scale.

CONCLUSION

The NS adequately identified the degree cutoffs of the overall severity of vocal deviation in the VAS with the linked speech task, as follows: no deviation – 0 to 27 mm, mild deviation – 27.1 to 50.5 mm, moderate deviation – 50.6 to 68.5 mm, intense deviation – 68.6 to 100 mm of the VAS. These results make it possible to compare the auditory-perceptual evaluation of the overall severity of vocal deviation between the VAS and the NS with the linked speech task. Further studies of the correspondence between these two scales should use both the linked speech and sustained vowel tasks to expand the use and applicability of the VAS in voice clinical practice and research.

ACKNOWLEDGMENTS

Gratitude is extended to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES for the grants provided: finance code 001 (PDPG do Programa de Pós-Graduação em Ciências Fonoaudiológicas) and process number 88887.615544/2021-00.

REFERENCES

1. Yamasaki R, Gama ACC. Desafios e referências na avaliação perceptivo-auditiva da voz. In: Thieme Revinter, Lopes L, Moreti F, Ribeiro LL, Pereira EC, organizadores. Fundamentos e atualidades em voz Clínica. Rio de Janeiro; 2019. p. 9-30.
2. Oliveira SB, Gama ACC, Chaves CR. Interference of background experience on agreement of perceptivo-auditory analysis of neutral and dysphonic voices. Distúrb. Comum. 2016;28(3):415-21. Disponível em: https://revistas.pucsp.br/index.php/dic/article/view/27706
3. Chan KMK, Yiu EML. A comparison of two perceptual voice evaluation training programs for naive listeners. J Voice. 2006;20(2):229-41. https://doi.org/10.1016/j.jvoice.2005.03.007 PMID: 16139475.
4. Behlau M, Rocha B, Englert M, Madazio G. Validation of the Brazilian Portuguese CAPE-V Instrument – Br CAPE-V for auditory-perceptual analysis. J Voice. 2022;36(4):586-91. https://doi.org/10.1016/j.jvoice.2020.07.007 PMID: 32811691.
5. Hirano M. Clinical examination of voice. Springer-Verlag. New York. 1981.
6. Simberg S, Laine A, Sala E, Rönnemaa A-M. Prevalence of voice disorders among future teachers. J Voice. 2000;14(2):231-5. https://doi.org/10.1016/S0892-1997(00)80030-2 PMID: 10875574.
7. Yamasaki R, Leão SHS, Madazio G, Padovani M, Azevedo R. Análise perceptivo-auditiva de vozes normais e alteradas: escala analógico visual. XV Congresso Brasileiro de Fonoaudiologia e VII Congresso Internacional de Fonoaudiologia; 2007 out 16-20; Gramado, Rio Grande do Sul, Brasil: Sociedade Brasileira de Fonoaudiologia. 2007.
8. Yamasaki R, Madazio G, Leão SHS, Padovani M, Azevedo R, Behlau M. Auditory-perceptual evaluation of normal and dysphonic voices using the Voice Deviation Scale. J Voice. 2017;31(1):67-71. https://doi.org/10.1016/j.jvoice.2016.01.004 PMID: 26873420.
9. Vieira MMRM, Yamasaki R, Brasolotto AG, Behlau M. Intensidade do desvio vocal na escala analógicovisual para adultos idosos. 21º Congresso Brasileiro e 2º Ibero-Americano de Fonoaudiologia; 22-25 set 2013; Porto de Galinhas, Recife, Brasil. Anais. São Paulo: Sociedade Brasileira de Fonoaudiologia; 2013. p. 542-6.
10. Contreras-Ruston F, Guzman M, Castillo-Allendes A, Cantor-Cutiva L, Behlau M. Auditory-perceptual assessment of healthy and disordered voices using the Voice Deviation Scale. J Voice. 2021;28(3):654-9. https://doi.org/10.1016/j.jvoice.2021.10.017 PMID: 34903393.
11. 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(3):279-84. https://doi.org/10.1590/2317-1782/20152014167 PMID: 26222946.
12. Baravieira PB, Brasolotto AG, Montagnoli NA, Silvério KCA, Yamasaki R, Behlau M. Auditory-perceptual evaluation of rough and breathy voices: Correspondence between analogical visual and numerical scale. CoDAS. 2016;28(2):163-7. https://doi.org/10.1590/2317-1782/20162015098 PMID: 27191880.
13. Kempester GB, Gerrat BR, Abbott KV, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. Am J Speech Lang Pathol. 2009;18(2):124-32. https://doi.org/10.1044/1058-0360(2008/08-0017 PMID: 18930908.
14. Brinca L, Batista AP, Tavares AI, Pinto PN, Araújo L. The effect of anchors and training on the reliability of voice quality ratings for different types of speech stimuli. J Voice. 2015;32(6):705-9. https://doi.org/10.1016/j.jvoice.2015.01.007 PMID: 25795348.
15. Lu FL, Matteson S. Speech tasks and interrater reliability in perceptual voice evaluation. J Voice. 2014;28(6):725-32. https://doi.org/1010.1016/j.jvoice.2014.01.018 PMID: 24841668.

16. Barsties B, De Bodt M. Assessment of voice quality: Current state-of-the-art. Auris Nasus Larynx. 2015;42(3):183-8. https://doi.org/1010.1016/j.anl.2014.11.001 PMID: 25440411.

17. Maryn Y, Roy N. Sustained vowels and continuous speech in the auditory-perceptual evaluation of dyspohonia severity. J Soc Bras Fonoaudiol. 2012;24(2):107-12. https://doi.org/10.1590/s2179-64912012000200003 PMID: 22832675.

18. Fleiss J. Statistical methods for rates and proportions. New York: John Wiley & Sons, 1981.

19. Chaves CR, Campbell M, Gama ACC. The influence of native language on auditory-perceptual evaluation of vocal samples completed by Brazilian and Canadian SLPs. J Voice. 2017;31(2):258.e1-258.e5. https://doi.org/10.1016/j.jvoice.2016.05.021 PMID: 27427162.

20. Miot HÁ. Análise de correlação em estudos clínicos e experimentais. J Vasc Bras. 2018;17(4)275-9. https://doi.org/10.1590/1677-5449.174118

Authors' contributions:

PCMS: Conceptualization; Investigation; Methodology; Visualization; Writing original draft; Writing - Review and editing.

ACCG: Data analysis; Funding acquisition; Methodology; Project administration; Supervision.

Data sharing statement:

We declare that the research data will not be shared.