

# GALP Qualifier Scale: Initial Considerations to Classify a Voice Problem

Marina Englert<sup>a, b</sup> Viviana Mendoza<sup>c</sup> Mara Behlau<sup>a, b</sup> Marc De Bodt<sup>c-e</sup>

<sup>a</sup>Human Communication Disorders, Universidade Federal de São Paulo UNIFESP, São Paulo, Brazil; <sup>b</sup>Centro de Estudos da Voz CEV, São Paulo, Brazil; <sup>c</sup>Department of Otorhinolaryngology, Head and Neck Surgery and Communication Disorders, University Hospital of Antwerp, Antwerp, Belgium; <sup>d</sup>Faculty of Medicine and Health Sciences, Antwerp University, Antwerp, Belgium; <sup>e</sup>Faculty of Medicine and Social Health Sciences, University of Ghent, Ghent, Belgium

## Keywords

Voice disorders · Dysphonia · Cluster analysis · International Classification of Functioning, Disability, and Health · European Laryngological Society · GALP scale

## Abstract

**Objective:** To propose a single qualifier scale for voice problems based on the International Classification of Functioning, Disability, and Health (ICF) that classifies a voice problem considering its multidimensionality. **Method:** A multicultural database was analyzed (280 subjects). The analyzed information was: the perceptual judgment of the overall voice quality (G); the acoustic analysis (A) with the Acoustic Voice Quality Index; the laryngeal diagnosis (L) and the patient self-assessment (P) using the Voice Handicap Index. The variables were categorized. A 2-step cluster analysis was performed to define groups with common characteristics. **Results:** A 7-point qualifier scale, the GALP, was defined to generally classify levels of voice problems considering 4 dimensions of the voice evaluation. Each level of voice problem, that is, no problem, mild, moderate, severe, or complete voice problem, has its own possible outcome for G, A, L, and P that will change, or not, the overall level of voice problem.

The extremes of the scale represent “no problem” at all when all parameters are normal, and “complete problem” when all parameters are altered. The 3 levels in between were defined by the cluster analysis (mild, moderate, and severe problem) and change according to the outcome of each evaluation (G, A, L, and P). Thus, changes in one parameter alone may or not contribute to the change of the level of voice problem. Also, there are 2 categories for cases that do not fit the classification (not specified) and for which some of the variables are missing (not applicable). **Conclusion:** The GALP scale was proposed to classify the level of voice problem. This approach considers important dimensions of voice evaluation according to the ICF. It is a potential tool to be used by different professionals, with different assessment procedures, and among different populations, clinicians, and study centers.

© 2019 S. Karger AG, Basel

## Introduction

The voice evaluation consists of the perceptual-auditory judgment, the acoustic analysis, the observation of the laryngeal structure, and the patient self-evaluation;

therefore, it is a multidimensional assessment. To guide the voice evaluation, the Committee on Phoniatics of the European Laryngological Society (ELS) proposed a protocol that evaluates the voice in 5 dimensions: (1) perceptual voice evaluation (grade, breathiness, and roughness), (2) acoustics, (3) videolaryngoscopy, (4) aerodynamics, and (5) subjective rating by the patient [1].

Most commonly, the protocols used to perceptually judge the voice quality are the GRBAS scale, a 4-point scale [2], and the visual analog scale, a 100-point scale. Both scales are highly correlated [2, 3] and present similar thresholds among different cultures [3–5]. This favors the use of the perceptual judgment cross-culturally to define the overall voice quality degree of deviation. However, this is a subjective assessment that suffers the influence of the rater's auditory experience, clinical experience, training, and cultural background [6–13]. For this reason, it must be complemented by objective measures such as the acoustic analysis [14, 15].

The acoustic analysis gives an objective value for the voice quality by analyzing the voice parameters. This analysis may be from an isolated parameter, such as jitter or shimmer, or the combination of parameters, using a multiparametric approach, such as the Dysphonia Severity Index, Acoustic Voice Quality Index (AVQI), and Cepstral Spectral Index of Dysphonia [16–19].

The AVQI is beginning to be widely used among different languages and cultures presenting good cross-cultural validity [20–28]. The index presents different thresholds among different countries to define whether the voice quality has deviation; these values vary from 1.33 to 2.43, considering the AVQI current version, 03.01 [23–27, 29]. Thus, the spoken language must be considered in this analysis.

The laryngeal assessment is also a very important aspect of voice evaluation; it is basic for voice diagnosis [30]. It provides information about the structure and gross function and measures of the vocal fold vibration [31]. The laryngeal assessment is usually the focus of medical examination [30] and provides an etiological laryngeal diagnosis that can be organic or functional [32]. These 2 categories can commonly be overlapped, once organic changes can generate functional disorders and vice versa [32]. The organic voice disorder is when there are changes in the larynx structure while in the functional voice disorder no changes are observed [32].

Finally, the patient's subjective rating is extremely important in voice evaluation. For this reason, many self-assessment protocols have been developed and validated in different languages. These protocols quantify the im-

port of a voice problem related to the quality of life (Voice-Related Quality of Life), the perceived handicap (Voice Handicap Index), and the restrictions in the activity and participation and disabilities (Voice Activity and Participation Profile) among others [31]. Thresholds that discriminate healthy from dysphonic individuals have been established; they vary among the different protocols but always present good sensitivity and specificity [33].

It is noteworthy that the ELS protocol was designed to observe the outcomes before and after therapy, and this can only be observed when the 5 dimensions of the voice assessment are considered; changes of parameters alone are weak, in other words, the improvement of one dimension alone has low redundancy to indicate the patient's improvement after therapy or any other intervention [34]. In this sense, the voice evaluation aims to assess the dysphonia in a more holistic way. In addition, the correlation of the perceptual auditory judgment and/or the acoustic analysis with the self-assessment protocols and/or the larynx diagnosis are not strong [19, 35–38]. Also, the scores of the self-assessment protocols are not influenced by an organic or functional dysphonia [36]. Nevertheless, despite the outcome for each dimension of the voice evaluation, the degree of dysphonia relies mainly on the degree of deviation of the voice quality; that is, no deviation, mild deviation, moderate deviation, or severe deviation [39, 40]. In other words, the degree of dysphonia is commonly given by the degree of the overall vocal deviation [16, 39].

It is strongly recommended by the World Health Organization (WHO) to consider other factors beyond the diagnosis when assessing a patient, understanding all health conditions [41]. Therefore, the WHO developed a framework for health and disability, the International Classification of Functioning, Disability, and Health, also known as ICF [42]. The ICF focuses on the people level of health instead of their disabilities. It has the intention to be complementary to the highly used International Statistical Classification of Diseases and Related Health Problems (ICD-10). The ICD-10 is basically used to classify causes of death, giving an etiological framework while the ICF considers the health conditions to classify functioning and disabilities.

There are 4 key concepts in the ICF: body functions, body structures, activities and participation, and contextual (environmental and personal) factors [42]. The “body functions” are related to the physiological function of the body system; it includes also psychological consequences, such as temperament and personality

functions and emotional functions. The “body structure” is related to the anatomical body parts. “Activities and participation” are related to the individual capacity of performing a task and how the person gets involved in it. “Contextual factors” are related to how the environment facilitates or not the person’s daily activities, that is, use of technological tools, workplace conditions, and government policies [42, 43].

The list of domains in ICF becomes a classification when qualifiers are used [44, 45]. The generic qualifiers are no problem (0), mild problem (1), moderate problem (2), severe problem (3), complete problem (4), not specified (8), and not applicable (9) [44, 45]. Hence, the ICF provides a framework aiming that people all around the world and from different fields can use common terms to communicate about health and health care. Thus, the ICF provides information that embraces all kinds of diseases and, therefore, all kinds of voice disorders.

The 4 key concepts of the ICF can be related to some of the dimensions of the voice evaluation as following: body functions are assessed by the perceptual voice judgment and acoustics analysis; the body structure is assessed by videolaryngoscopy; the activities and participation are assessed by means of self-assessment protocols [43]. The contextual concept, that is the environmental and personal factors, can be assessed by the conversation with the patient while understanding his case history [43].

The body function is highly considered to determine the functioning level of a condition [46, 47]. On the other hand, contextual factors seem to be the least considered [48]. One important aspect of clinical practice is to understand the patient case history. However, there is no specific tool to assess environmental factors related to voice, besides an informal conversation and isolated questions of self-assessment protocols. Nevertheless, information gathered by talking to the patient in a semistructured interview and then using the ICF to identify and qualify the information of the patient’s functioning can be successfully done [46, 49].

The different dimensions of voice evaluation are necessary to properly classify a voice problem. However, there is not a classification system that defines the overall degree of a voice problem considering all dimensions of the voice assessment. Therefore, the aim of this study is to propose a concept of a single qualifier scale, based on the ICF, that classifies a voice problem considering its multidimensionality including information of the perceptual, acoustic, and laryngeal assessment, and the patient self-evaluation.

## Materials and Methods

This study was performed using the databases from 2 different countries with information gathered during the routine clinical voice assessment. Data from 280 speakers (150 Brazilian Portuguese and 130 Belgian Dutch speakers) were analyzed. The speaker’s age ranged from 18 to 89 years (115 men and 165 women, mean of 47.91 and SD of 19.15 years). The information corresponding to each speaker considered was: the G score, the AVQI score, the Voice Handicap Index (VHI) or the VHI-10 total score, and the laryngeal diagnosis. Patients from the database that had incomplete information were excluded.

The G score is the overall voice quality from the GRBAS scale [2]. It classifies the voice quality in a 4-point scale in which 0 = no vocal deviation, 1 = mild deviation, 2 = moderate deviation, and 3 = severe deviation. The G was previously judged by the voice specialist at the moment of the patient assessment. The voice specialist from each country had over 10 years of clinical experience and intrarater reliability ranging from 0.605 to 0.773 (Cohen’s kappa).

The recording of the Belgian database was previously performed using the AKG C555L head-mounted condenser microphone, connected to the external hardware Computerized Speech Lab Model 4500 (CSL, KayPENTAX Corp.). The recordings of the Brazilian database were performed using the AKG C420 head-mounted condenser microphone, the AKG MPA V L cable, and the Focusrite iTrack Solo audio interface connected to a PC with the Audacity program (version 2.0.6). Samples from both databases were digitized at a rate of 44.1 kHz, resolution of 16 bits, and saves as .wav. All recordings took place in a quiet room.

The AVQI is a multiparametric acoustic index that runs as a Praat script, a freeware software for speech analysis in phonetics. The AVQI provides one score for the overall voice quality considering 6 acoustic measurements: the smoothed cepstral peak prominence, harmonics-to-noise ratio, shimmer percent, shimmer decibel, general slope of the spectrum, and tilt of the regression line through the spectrum [18]. The score ranges from 0 to 10 and considers continuous speech and sustained phonation [18]. It has been validated in different languages demonstrating high diagnostic accuracy and presenting varied threshold values to identify voices with and without vocal deviation. In addition, there is more than one AVQI version (first version, 02.01 and 03.01), and each of its versions also presents different thresholds [20–28].

Laryngeal diagnosis is an essential parameter of the voice assessment providing important information of the larynx structure that will guide the treatment. The diagnosis is based on the laryngeal endoscopic imaging, videoendoscopy (that uses a constant light source to assess the larynx structures and functions) and videostroboscopy (that relies on the strobe effect and assesses the vocal folds’ vibratory function) [31]. Commonly voice disorders can be organic or functional [32].

The VHI is a self-assessment protocol translated and validated in many languages. It has a short version with 10 statements, VHI-10, and a complete version with 30 statements, VHI. The VHI has high accuracy in detecting perceived voice impairment with an established cutoff point; these cutoff points vary for each of its versions [33].

In order to standardize the outcomes of the clinical measurements, the variables were categorized as follows:

**Table 1.** Frequency distribution of the categorized variables

	BR Portuguese, n (%)	BE Dutch, n (%)	BR and BE, n (%)
<b>G</b>			
No deviation	32 (21.3)	32 (24.62)	64 (22.86)
Mild	59 (39.3)	41 (31.54)	100 (35.71)
Moderate	40 (26.6)	36 (27.69)	76 (27.14)
Severe	19 (12.6)	21 (16.15)	40 (14.29)
<b>AVQI</b>			
Below threshold	55 (36.6)	26 (20)	81 (28.93)
Above threshold	95 (63.3)	104 (80)	199 (71.07)
<b>Larynx</b>			
Normal	90 (60)	63 (48.46)	153 (54.64)
Altered	60 (40)	67 (51.54)	127 (45.36)
<b>Self-assessment</b>			
Below cutoff point	73 (48.6)	52 (40)	125 (44.64)
Above cutoff point	77 (51.3)	78 (60)	155 (55.36)

BR, Brazilian; BE, Belgian; G, perceptual judgment of the overall voice quality; AVQI, Acoustic Voice Quality Index.

G score (G): no deviation, mild deviation, moderate deviation, or severe deviation;

AVQI score (A): below or above the threshold, considering the threshold for each AVQI version and language, that was 1.33 for the Brazilian database (AVQI version 03.01) and 2.95 for the Dutch database (AVQI first version);

larynx (L): no visible alteration observed by laryngoscopy (such as functional dysphonia) or presence of visible alteration observed by laryngoscopy (such as nodules, vocal fold paralysis, cyst). The larynx diagnosis was provided by an otolaryngologist using videoendoscopy recordings. The medical laryngeal diagnosis of both databases was considered for the study.

VHI score (P): below or above the cutoff point for each version of the protocol, that was 20 for the VHI and 8 for the VHI-10.

Table 1 presents the frequency distribution of each categorized variable.

### Statistical Analysis

A cluster analysis was performed in order to define natural clusters within the data sets that are not evident in classifying a voice problem. The SPSS version 25 was used as the statistical program to execute 2-step cluster analysis. This analysis can be used for categorical variables with 3 or more attributes [50]. The silhouette measure of cohesion and separation was used to evaluate the clustering solution's overall goodness of fit [51].

## Results

A qualifier scale with 7 generic categories was established: no voice problem (0), mild voice problem (1), moderate voice problem (2), severe voice problem (3),

complete voice problem (4), not specified (8), and not applicable (9). The extremes, that is no problem and complete problem, were predefined by the authors. There are also 2 other categories: not specified, for cases where a person does not fit within any category, and not applicable, when there is a lack of information not allowing to place the person in the GALP scale.

The empirical classification of the other 3 distinctive groups (mild, moderate, and severe voice problem) was the result of the 2-step cluster analysis. The averaged silhouette coefficient was equal to 0.5, indicating a fair solution for the overall goodness of fit of the clustering solutions.

The variable distributions in each cluster were consistent among both databases. Figure 1 presents the relative distribution of the variables in the clusters for each database separately and both together. The predictor importance, in grayscale, indicates the importance of each variable in making a prediction. In the present study, the variables were organized in descending order of importance as AVQI, VHI, G, and larynx. The histograms show the relative frequency distribution of the categorized variables among the 3 clusters. Table 2 complements this information with the values of the observed frequency distribution; from these statistics it can be inferred the most or least frequent parameter in each degree of voice problem. Additionally, the size of each cluster represents the total number of observations contained within a category. It can be observed that the 3 different categories have approximately the same percentage among the different databases.

According to the qualifier scale, each cluster was interpreted considering the perceptual judgment (G), the acoustic analysis (A), the laryngeal observation (L) and the patient self-assessment (P) (Table 3).

## Discussion

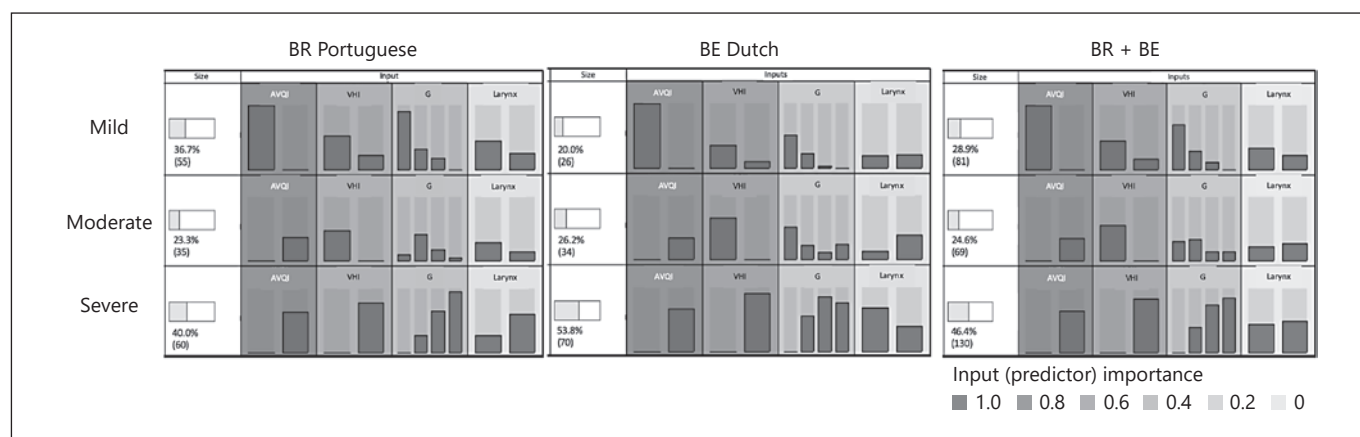
The degree of voice disorder, despite the protocol used for its assessment, is mainly classified focused on evaluator judgments of the voice quality [16]. The proposed GALP scale classifies a voice problem considering 4 dimensions of the voice evaluation and plotting them in a generic qualifier scale based on the ICF [42]. Therefore, traditional voice assessment is implemented according to the ICF guidelines.

The GALP scale is composed of 7 different categories to classify a voice problem considering the perceptual overall voice quality judgment (G), the acoustic analysis

**Table 2.** Frequency distribution, in percentage, of the variables in the clusters for the Brazilian Portuguese, Belgian Dutch and both databases

	Mild problem			Moderate problem			Severe problem		
	BR Portuguese	BE Dutch	BR and BE	BR Portuguese	BE Dutch	BR and BE	BR Portuguese	BE Dutch	BR and BE
G									
No deviation	90.6	50	70.3	9.4	50	29.7	0	0	0
Mild	32.2	22	28	40.7	22	33	27.1	56.1	39
Moderate	17.5	2.8	10.5	17.5	11.1	14.5	65	86.1	75
Severe	0	0	0	5.3	23.8	15	94.7	76.2	85
AVQI									
Below threshold	100	100	100	0	0	0	0	0	0
Above threshold	0	0	0	36.8	32.7	34.7	63.2	67.3	65.3
Larynx									
Normal	44.4	19	34	28.9	12.7	22.2	26.7	68.3	43.8
Altered	25	20.9	22.8	15	38.8	27.6	60	40.3	49.6
Self-assessment									
Below cutoff point	52.1	34.6	44.8	47.9	65.4	55.2	0	0	0
Above cutoff point	22.1	10.3	16.1	0	0	0	77.9	89.7	83.9

BR, Brazilian; BE, Belgian; G, perceptual judgment of the overall voice quality; AVQI, Acoustic Voice Quality Index.



**Fig. 1.** Relative distribution of the variables in the clusters for the Brazilian (BR) Portuguese, Belgium (BE) Dutch and both databases.

(A), the laryngeal imaging observation (L), and patient self-assessment (P), providing a global overview of the voice problem. To better introduce the new scale and how it should be used, a detailed description of each category of the GALP will be given as follows.

“No problem”: includes G equal to no deviation or mild deviation. The A must be below the threshold established for the language that is being assessed and the acoustic instrument that is being considered. The L must be normal, no visible laryngeal alteration observed by la-

ryngoscopy. The P must be below the cutoff point established for the instrument that is being used, indicating that self-assessment of the voice impact is within normal thresholds.

“Mild problem”: includes everything that has been specified in the previous category plus G can present moderate deviation, L can also be altered (it is possible to have structural laryngeal alteration), and P can be above the cutoff point. It is possible that the 4 aspects from the “mild problem” coincide with the “no problem”; in this

**Table 3.** Presentation of the GALP

	G	A	L	P
No problem – 0	no and mild	below AVQI threshold	normal	below VHI cutoff point
Mild problem – 1	no, mild, and moderate	below AVQI threshold	normal or altered	below or above VHI cutoff point
Moderate problem – 2	no, mild, moderate, and severe	above AVQI threshold	normal or altered	below VHI cutoff point
Severe problem – 3	mild, moderate, and severe	above AVQI threshold	normal or altered	above VHI cutoff point
Complete problem – 4	moderate and severe	above AVQI threshold	altered	above VHI cutoff point
Not specified – 8	information does not fit into the GALP			
Not applicable – 9	lack of clinical information			

AVQI, Acoustic Voice Quality Index; VHI, Voice Handicap Index.

case, another aspect of the ICF should be considered, the contextual factor. The contextual factors are evaluated by means of conversation with the patient [46, 49]. Therefore, the clinician should decide in which category the patient should be placed, taking into account arguments related to the contextual factors.

“Moderate problem”: the G plays a minor role once all perceived vocal judgments are possible, from no deviation to severe deviation. The A must be above the threshold established for the language that is being assessed and the acoustic instrument that is being considered; the L can or not present structural laryngeal alteration. What is interesting in this category is that P must be below the cutoff point, therefore, the patient may have bad voice quality and laryngeal alteration but no impact in his self-assessment. This is highly important in this category because it will be basically the patient self-assessment that will position him as having “moderate problem” or “severe problem.”

“Severe problem”: includes all G except no deviation, A must be above the threshold and L may or may not have visible laryngeal alteration. Differently from the previous category, P must be above the cutoff point, this means that the patient has a self-perceived vocal impairment.

“Complete problem”: this category includes the worst cases in clinical practice, G equal to moderate or severe deviation, A above the threshold, L with visible laryngeal alteration and P above the cutoff point. It is possible that the 4 aspects from this category coincide with the severe problem, equally to “no problem” and “mild problem.” Similarly, the contextual factors must be considered for the clinician to gather arguments to decide whether the patient has “severe” or “complete voice problem.”

“Not specified” and “not applicable”: these 2 categories must be considered for cases that do not fit into the GALP. One example is the perceptual judgment of severe vocal

deviation, that is, G3, AVQI below the threshold, larynx with visible lesion and VHI score above the cutoff point. The main inconsistency in this example is the G3 and the AVQI below the threshold; if the acoustic analysis was above the threshold, this example would be placed at the “severe problem.” The acoustic analysis is strongly correlated with the perceptual analysis [16]. However, as expected, this is not a perfect correlation, and cases where the perceptual judgment does not match with the acoustic analysis are possible [52]. Therefore, cases such as this should be classified as “not specified.” If there is no clinical information for one of the voice evaluation dimensions considered in the GALP, the case should be classified as “not applicable,” once there are not enough data to place the patient within the GALP.

The extremes of the scale were established by the authors based on previous studies and experience. According to the WHO “no problem” is defined as the absence of impairment and “complete problem” is a total problem [42]. Thus, for 1 subject to be classified as “no voice problem” all the assessed dimensions, that is, G, A, L, and P, must be normal. Otherwise, if 1 case is classified as “complete problem,” all the assessed dimensions, that is, G, A, L, and P, must be altered. In addition, the voice evaluation is more reliable, and raters agree more consistently to assess the extremes, that is, normal and extreme dysphonic voices [6, 53–56]. Therefore, it is not difficult in clinical practice to classify a person with no voice problem or with complete problem; on the other hand, the intermediate categories are harder to be defined, and this is exactly where the GALP may contribute.

There are different thresholds standardized worldwide for each dimension of voice assessment, which makes comparisons of dysphonia outcomes harder among different countries [20–27]. The GALP organized a first attempt to qualify a voice problem considering the G score,

reliable worldwide [3–5] the laryngeal diagnosis as with or without visible alteration observed by laryngoscopy, the threshold of the acoustic analysis and the cutoff point of a self-assessment protocol.

Ideally, if the acoustic analysis procedure and the subjective rating by the patient are performed with a previously validated procedure or protocol with defined thresholds, the GALP can be applied. Hence, comparison among different cultures that use different tools and protocols for the voice assessment may be performed in a more reliable and robust way once there will be a unified scale. To verify this hypothesis multicultural studies involving different acoustic tools and self-assessment protocols should be performed.

The GALP considers almost all ELS outcome measures. This is in accordance with the WHO recommendation of considering the patient's health and not only disabilities [41, 57]. Previous reports of related fields have also worked on developing protocols that assess the patient's health considering the ICF [57, 58]. However, these studies classified each isolated item of the ICF regarding a specific health condition and not grouping them to provide an overall category for the problem.

It is noteworthy that the ICF does not give a global classification for somebody's health, it classifies many domains related to body structures, body function, activity and participation, and environmental factors providing a code for each one based on the qualifier scale [59]. The GALP, on the other hand, aims to summarize the outcomes of these different domains giving an overall classification for a voice problem.

The GALP presents an initial concept to potentially classify a voice problem considering 3 key concepts of the ICF combined with 4 dimensions of the ELS: body function (perceptual judgment of the voice quality and acoustic analysis), body structure (laryngeal assessment), and the activities and participation (self-assessment protocols). Further studies are needed to assess the GALP responsiveness, reliability, stability, and sensitivity. Also, these studies should compare the GALP outcome among different cultures, different voice assessment tools for the acoustic analysis, and different self-assessment protocols, and its outcomes before and after treatment must be analyzed. Therefore, the GALP reliability and robustness will be assessed and may guarantee its use in the voice clinical routine and research.

#### *Study Limitations*

The use of thresholds was decided in order to better adapt the usage of different instruments and protocols to

classify the patient's voice problem. Hence, improvements within a category can be neglected; for example, a VHI score from 100 to 21 and an AVQI score from 9.0 to 3.0 are information that will not be sensible using the GALP.

It would have been interesting to use the Voice Activity and Participation Profile (VAPP) self-assessment tool in the GALP analysis, once it was developed in accordance with activities and participation concepts. The answers of the VAPP are given in a visual analog scale, which is not as easy to use as an equal-appearing interval scaling [60]; however, it is a reliable self-assessment protocol with defined cutoff points for each concept that it assesses, activity limitations, and participation restriction [61]. Considering that the GALP scale can be used with any self-assessment protocol that has a defined cutoff point, further studies using the GALP and the VAPP can be developed; therefore, this is also a strength of the GALP scale. The database of the present research had information of the self-assessment protocol VHI; thus, the VHI was considered for the initial consideration of the GALP.

## **Conclusion**

A scale to classify the level of voice problem was proposed and named the GALP. This approach considers important dimensions of the voice evaluation according to the ELS and the ICF. The GALP scale seems to be a potential tool to be used by different professionals and with different acoustic measurements and self-assessment protocols, as long as a defined threshold exists for each of these instruments. In addition, the GALP seems to be a robust and promising tool to standardize the level of voice problem among different populations, clinicians and study centers.

## **Acknowledgment**

This work was partially supported by a grant from the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie (grant agreement No. 766287).

## **Statement of Ethics**

The study followed the principles of the Declaration of Helsinki.

## Disclosure Statement

The authors have no conflicts of interest to declare.

## Funding Sources

No funding sources were used for the research.

## Author Contributions

M.E. was responsible for the study design, data analysis, and writing of the manuscript; V.M. was responsible for the study design, data analysis, statistical analysis, and writing of the manuscript; M.B. and M.B. were responsible for the study design, data analysis, and writing and revision of the manuscript.

## References

- 1 Friedrich G, Dejonckere PH. [The voice evaluation protocol of the European Laryngological Society (ELS)—first results of a multicenter study]. *Laryngorhinootologie*. 2005 Oct;84(10):744–52.
- 2 Hirano M. *Clinical examination of voice*. New York (NY): Springer Verlag; 1981.
- 3 Lee YW, Kim GH, Bae IH, Park HJ, Wang SG, Kwon SB. The cut-off analysis using visual analogue scale and cepstral assessments on severity of voice disorder. *Logoped Phoniater Vocol*. 2018 Dec;43(4):175–80.
- 4 Simberg S, Laine A, Sala E, Rönnemaa AM. Prevalence of voice disorders among future teachers. *J Voice*. 2000 Jun;14(2):231–5.
- 5 Yamasaki R, Madazio G, Leão SH, Padovani M, Azevedo R, Behlau M. Auditory-perceptual Evaluation of Normal and Dysphonic Voices Using the Voice Deviation Scale. *J Voice*. 2017 Jan;31(1):67–71.
- 6 Gerratt BR, Kreiman J, Antonanzas-Barroso N, Berke GS. Comparing internal and external standards in voice quality judgments. *J Speech Hear Res*. 1993 Feb;36(1):14–20.
- 7 Yiu EM, Murdoch B, Hird K, Lau P. Perception of synthesized voice quality in connected speech by Cantonese speakers. *J Acoust Soc Am*. 2002 Sep;112(3 Pt 1):1091–101.
- 8 Eadie TL, Baylor CR. The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *J Voice*. 2006 Dec;20(4):527–44.
- 9 Yiu EM, Murdoch B, Hird K, Lau P, Ho EM. Cultural and language differences in voice quality perception: a preliminary investigation using synthesized signals. *Folia Phoniater Logop*. 2008;60(3):107–19.
- 10 Sellars C, Stanton AE, McConnachie A, Dunnet CP, Chapman LM, Bucknall CE, et al. Reliability of perceptions of voice quality: evidence from a problem asthma clinic population. *J Laryngol Otol*. 2009 Jul;123(7):755–63.
- 11 Eadie TL, Kapsner-Smith M. The effect of listener experience and anchors on judgments of dysphonia. *J Speech Lang Hear Res*. 2011 Apr;54(2):430–47.
- 12 Schaeffer N. Student training to perceptually assess severity of dysphonia using the dysphonic severity percentage scale. *J Voice*. 2013 Sep;27(5):611–6.
- 13 Englert M, Madazio G, Gielow I, Lucero J, Behlau M. Learning factor influence on the perceptual-auditory analysis. *Codas*. 2018 Jun;30(3):e20170107.
- 14 Roy N, Mazin A, Awan SN. Automated acoustic analysis of task dependency in adductor spasmodic dysphonia versus muscle tension dysphonia. *Laryngoscope*. 2014 Mar;124(3):718–24.
- 15 Barsties B, De Bodt M. Assessment of voice quality: current state-of-the-art. *Auris Nasus Larynx*. 2015 Jun;42(3):183–8.
- 16 Roy N, Barkmeier-Kraemer J, Eadie T, Sivasankar MP, Mehta D, Paul D, Hillman R. Evidence-based clinical voice assessment: a systematic review. *Am J Speech Lang Pathol*. 2013 May;22(2):212–26.
- 17 Wuyts FL, De Bodt MS, Molenberghs G, Remacle M, Heylen L, Millet B, et al. The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach. *J Speech Lang Hear Res*. 2000 Jun;43(3):796–809.
- 18 Maryn Y, De Bodt M, Roy N. The Acoustic Voice Quality Index: toward improved treatment outcomes assessment in voice disorders. *J Commun Disord*. 2010 May-Jun;43(3):161–74.
- 19 Awan SN, Roy N, Zhang D, Cohen SM. Validation of the Cepstral Spectral Index of Dysphonia (CSID) as a Screening Tool for Voice Disorders: Development of Clinical Cutoff Scores. *J Voice*. 2016 Mar;30(2):130–44.
- 20 Reynolds V, Buckland A, Bailey J, Lipscombe J, Nathan E, Vijayasekaran S, et al. Objective assessment of pediatric voice disorders with the acoustic voice quality index. *J Voice*. 2012 Sep;26(5):672.e1–7.
- 21 Maryn Y, De Bodt M, Barsties B, Roy N. The value of the acoustic voice quality index as a measure of dysphonia severity in subjects speaking different languages. *Eur Arch Otorhinolaryngol*. 2014 Jun;271(6):1609–19.
- 22 Uloza V, Petrauskas T, Padervinskis E, Ulozaitė N, Barsties B, Maryn Y. Validation of the Acoustic Voice Quality Index in the Lithuanian Language. *J Voice*. 2017 Mar;31(2):257.e1–11.
- 23 Barsties B, Maryn Y. External Validation of the Acoustic Voice Quality Index Version 03.01 With Extended Representativity. *Ann Otol Rhinol Laryngol*. 2016 Jul;125(7):571–83.
- 24 Hosokawa K, Barsties V, Latoszek B, Iwahashi T, Iwahashi M, Iwaki S, et al. The Acoustic Voice Quality Index Version 03.01 for the Japanese-speaking Population. *J Voice*. 2019 Jan;33(1):125.e1–125.e12
- 25 Barsties V, Latoszek B, Lehnert B, Janotte B. Validation of the Acoustic Voice Quality Index Version 03.01 and Acoustic Breathiness Index in German. *J Voice*. 2018 Sep;pii: S0892-1997(18)30231-5.
- 26 Delgado Hernández J, León Gómez NM, Jiménez A, Izquierdo LM, Barsties V, Latoszek B. Validation of the Acoustic Voice Quality Index Version 03.01 and the Acoustic Breathiness Index in the Spanish language. *Ann Otol Rhinol Laryngol*. 2018 May;127(5):317–26.
- 27 Pommée T, Maryn Y, Finck C, Morsomme D. Validation of the Acoustic Voice Quality Index, Version 03.01, in French. *J Voice*. 2018 Dec;pii:S0892-1997(18)30517-4.
- 28 Englert M, Lima L, Constantini AC, Latoszek BB, Maryn Y, Behlau M. Acoustic Voice Quality Index - AVQI for Brazilian Portuguese speakers: analysis of different speech material. *CoDAS*. 2019 Feb;31(1):e20180082.
- 29 Englert M, Barsties B, Maryn Y, Behlau M. Validation of the Acoustic Voice Quality Index, version 03.01, to the Brazilian Portuguese Language. [Under review].
- 30 Behlau M, Madazio G, Oliveira G. Functional dysphonia: strategies to improve patient outcomes. *Patient Relat Outcome Meas*. 2015 Dec;6:243–53.
- 31 Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyski D, Eadie T, et al. Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *Am J Speech Lang Pathol*. 2018 Aug;27(3):887–905.
- 32 Milutinović Z. Classification of voice pathology. *Folia Phoniater Logop*. 1996;48(6):301–8.
- 33 Behlau M, Madazio G, Moreti F, Oliveira G, Dos Santos LM, Paulinelli BR, et al. Efficiency and Cutoff Values of Self-Assessment Instruments on the Impact of a Voice Problem. *J Voice*. 2016 Jul;30(4):506.e9–18.



- 34 DeJonckere PH, Crevier-Buchman L, Marie JP, Moerman M, Remacle M, Woisard V; European Research Group on the Larynx. Implementation of the European Laryngological Society (ELS) basic protocol for assessing voice treatment effect. *Rev Laryngol Otol Rhinol (Bord)*. 2003;124(5):279–83.
- 35 Bassi IB, Assunção AA, de Medeiros AM, de Menezes LN, Teixeira LC, Côrtes Gama AC. Quality of life, self-perceived dysphonia, and diagnosed dysphonia through clinical tests in teachers. *J Voice*. 2011 Mar;25(2):192–201.
- 36 Behlau M, Zambon F, Moreti F, Oliveira G, de Barros Couto E Jr. Voice Self-assessment Protocols: Different Trends Among Organic and Behavioral Dysphonias. *J Voice*. 2017 Jan;31(1):112.e13–27.
- 37 Spina AL, Crespo AN. Assessment of Grade of Dysphonia and Correlation With Quality of Life Protocol. *J Voice*. 2017 Mar;31(2):243.e21–6.
- 38 Pommee T, Maryn Y, Finck C, Morsomme D. The Acoustic Voice Quality Index, Version 03.01, in French and the Voice Handicap Index. *J Voice*. 2018 Dec;pii:S0892-1997(18)30342-4.
- 39 Baken RJ. Chapter 1: Introduction. Clinical Measurement of Speech and Voice. Little Brown and Company (Inc.). British Library Cataloguing in Publication Data, Printed in the United States of America; 1987. pp. 1–5. ISBN: 0-85066-651-1.
- 40 Awan SN, Roy N. Toward the development of an objective index of dysphonia severity: a four-factor acoustic model. *Clin Linguist Phon*. 2006 Jan-Feb;20(1):35–49.
- 41 Botha A, Ras E, Abdoola S, Van der Linde J. Dysphonia in adults with developmental stuttering: A descriptive study. *S Afr J Commun Disord*. 2017 Jun;64(1):e1–7.
- 42 World Health Organization. International Classification of Functioning, Disability and Health (ICF). Geneva, Switzerland: WHO; 2001.
- 43 Ma EP, Yiu EM, Abbott KV. Application of the ICF in voice disorders. *Semin Speech Lang*. 2007 Nov;28(4):343–50.
- 44 World Health Organization. *Towards a Common Language for Functioning*. Geneva: Disability and Health; 2002.
- 45 World Health Organization. The ICF checklist version 2.1a. Clinical form for International Classification of Functioning, Disability, and Health. Geneva: WHO; 2002.
- 46 Tschiesner U, Linseisen E, Baumann S, Siedek V, Stelter K, Berghaus A, et al. Assessment of functioning in patients with head and neck cancer according to the International Classification of Functioning, Disability, and Health (ICF): a multicenter study. *Laryngoscope*. 2009 May;119(5):915–23.
- 47 Rumbach A, Aiken P, Novakovic D. Outcome Measurement in the Treatment of Spasmodic Dysphonia: A Systematic Review of the Literature. *J Voice*. 2018 Apr;pii:S0892-1997(18)30077-8.
- 48 Jacob M, Cox SR. Examining transgender health through the International Classification of Functioning, Disability, and Health's (ICF) Contextual Factors. *Qual Life Res*. 2017 Dec;26(12):3177–85.
- 49 Ávila CC, Cieza A, Anaya C, Ayuso-Mateos JL. The patients' perspective on relevant areas and problems in the bipolar spectrum disorder: individual interviews using the International Classification of Functioning, Disability and Health as a reference tool. *Am J Phys Med Rehabil*. 2012 Feb;91(13 Suppl 1):S181–8.
- 50 IBM. TwoStep Cluster Analysis [Internet]. United States of America: IBM; 2013 [2013]. Available from: [https://www.ibm.com/support/knowledgecenter/en/SSLVMB\\_22.0.0/com.ibm.spss.statistics.help/spss/base/idh\\_twestep\\_main.htm](https://www.ibm.com/support/knowledgecenter/en/SSLVMB_22.0.0/com.ibm.spss.statistics.help/spss/base/idh_twestep_main.htm)
- 51 Goodness Measures IB. *cluster evaluation algorithms* [Internet] United States of America: IBM; 2013. Available from [https://www.ibm.com/support/knowledgecenter/SSLVMB\\_22.0.0/com.ibm.spss.statistics.algorithms/alg\\_cluster-evaluation\\_goodness.htm](https://www.ibm.com/support/knowledgecenter/SSLVMB_22.0.0/com.ibm.spss.statistics.algorithms/alg_cluster-evaluation_goodness.htm)
- 52 Kent R. Hearing and Believing: Some Limits to the Auditory-Perceptual Assessment of Speech and Voice Disorders. *Am J Speech Lang Pathol*. 1996;5(3):7–23.
- 53 Kreiman J, Gerratt BR. Validity of rating scale measures of voice quality. *J Acoust Soc Am*. 1998 Sep;104(3 Pt 1):1598–608.
- 54 Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen N, Wilson JA. The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Otorhinolaryngol*. 2004 Sep;261(8):429–34.
- 55 Bele IV. Reliability in perceptual analysis of voice quality. *J Voice*. 2005 Dec;19(4):555–73.
- 56 Kreiman J, Gerratt BR. Comparing two methods for reducing variability in voice quality measurements. *J Speech Lang Hear Res*. 2011 Jun;54(3):803–12.
- 57 Kisser U, Adderson-Kisser C, Coenen M, Stier-Jarmer M, Becker S, Sabariego C, et al. The development of an ICF-based clinical guideline and screening tool for the standardized assessment and evaluation of functioning after head and neck cancer treatment. *Eur Arch Otorhinolaryngol*. 2017 Feb;274(2):1035–43.
- 58 Nawka T, Rosanowski F, Gross M. [How to render an expert opinion on dysphonia]. *Laryngorhinootologie*. 2014 Sep;93(9):591–8.
- 59 WHO. ICF e-learning Tool. Ludwig-Maximilians-Universität (LMU) Munich; 2018 [2018] Available from: <https://www.icf-elearning.com>
- 60 Ma EP, Yiu EM. Scaling voice activity limitation and participation restriction in dysphonic individuals. *Folia Phoniatr Logop*. 2007;59(2):74–82.
- 61 Lee SJ, Choi HS, Kim H, Byeon HK, Lim SE, Yang MK. Korean version of the Voice Activity and Participation Profile (K-VAPP): a validation study. *Commun Sci Disord*. 2016;21(4):695–708.