Efficiency and Cutoff Values of Self-Assessment Instruments on the Impact of a Voice Problem

*,†Mara Behlau, *Glaucya Madazio, *,†Felipe Moreti, *Gisele Oliveira, *Luciana de Moraes Alves dos Santos, *Bruna Rabelo Paulinelli, and *Euro de Barros Couto Junior, *†São Paulo, São Paulo, Brazil

Summary: Objectives. To evaluate the efficiency of four self-assessment questionnaires that rate the impact of a voice problem on the individual's life: Voice-Related Quality of Life (V-RQOL), the original and reduced versions of the Voice Handicap Index (VHI) and VHI-10, Vocal Performance Questionnaire (VPQ), and Voice Symptom Scale (VoiSS).

Methods. Data from 975 subjects, 486 with a diagnosis of dysphonia and 489 vocally healthy individuals, were submitted to the receiver operating characteristic (ROC) curve analysis to obtain the cutoff values that determine the discriminating power of these instruments (presence of dysphonia vs healthy voice).

Results. The ROC curve analysis showed that the most efficient questionnaires were the VoiSS and the VHI. Results showed that they presented as a perfect classification based on their efficiency, specificity, and sensitivity values (all three of them = 1). The VHI-10 and the V-RQOL showed excellent classification (VHI-10: efficiency = 0.991; specificity = 1; sensitivity = 0.981; V-RQOL: efficiency = 0.914; specificity = 0.860; sensitivity = 0.967). Finally, the VPQ showed a good level of classification (efficiency = 0.828; specificity = 0.824; sensitivity = 0.831). The cutoff values for the instruments are as follows: VoiSS = 16 points, VHI = 19 points, VHI-10 = 7.5 points, V-RQOL = 91.25, and VPQ = 20.5 points. These values are important for screening large populations as well as for helping in the decision-making process of clinical management. The cutoff values for maximum sensitivity and specificity of the instruments that did not produce perfect classification are as follows: VHI-10: sensitivity = 5; specificity = 7.5, V-RQOL: sensitivity = 86.25; specificity = 98.75, and VPQ: sensitivity = 15.5; specificity = 31.5.

Conclusions. Both the VoiSS and the VHI are perfect classifiers. The VHI-10 and the V-RQOL are excellent classifiers, and the VPQ is good at discriminating individuals with dysphonia from the ones without dysphonia.

Key Words: Voice–Dysphonia–Self-assessment–Validation studies–Protocols–Speech, Language, and Hearing Sciences.

INTRODUCTION

Voice disorders, also called dysphonias, occur in 3–9% of the population.[1] They affect quality of life in several different ways. Traditionally, the evaluation of patients with voice disorders is a multidimensional process, including at least a laryngeal examination, perceptual, and acoustic analysis.[2] However, the patient's experience of living with dysphonia cannot be inferred directly by these standard clinical assessments. Measuring what patients perceive about their health condition is essential, especially because there is typically a low correlation between the patient's and the clinician's subjective voice analyses.[3,4] Therefore, only the patient can provide real information about his/her experience with the voice problem, which cannot be obtained with any objective analysis.[5,6]

The concept of health evolved remarkably during the last decades after the formulation of the International Classification of Functioning.[7] The concept advanced toward measuring a disability as a difficulty found at least in one of the three functioning domains: impairment, activity limitations, and participation restriction. Disability is a consequence of the interaction between health and contextual factors, whether the latter are environmental or personal. Hence, it is essential for the diagnostic process to understand the perspective of the individual who experiences the problem. The perspective of the patient is usually obtained by means of self-assessment questionnaires that rate the impact of a certain deviation, disorder, or illness.

Numerous self-assessment questionnaires have been developed since the 1990s.[8–12] They were developed not only to quantify the impact of a voice problem and to evaluate the patient's progress but also to contribute in therapeutic management. These instruments became very popular both in the clinical and scientific settings. They achieved fast international popularity never seen before with any other approach, including acoustic analysis.[13] The Voice Handicap Index (VHI)[8] was the first questionnaire introduced in the area. It has a specific purpose of assessing the impact of dysphonia on patients' quality of life. This totally new perspective was proposed during a time when the trend was to use quantitative measures from heavy instrumentation for vocal analysis. Although the several other self-assessment instruments that followed the VHI were built in a somewhat structured manner, their development was based on varied criteria. This is specifically true in regard to the conceptual and empirical basis used for generating the instruments' content, that is, compiling of items. Some instruments included only data from patients' records, others included health professionals' and patients' points of view and/or the scientific literature in the area.

Accepted for publication May 11, 2015.

From the *Centro de Estudos da Voz–CEV, São Paulo, São Paulo, Brazil; and the †Department of Speech-Language Pathology and Audiology, Universidade Federal de São Paulo–UNIFESP, São Paulo, São Paulo, Brazil.

Address correspondence and reprint requests to Mara Behlau, Rua Machado Bittencourt, 361. São Paulo, SP 04044-001, Brazil. E-mail: mbehlau@uol.com.br

Journal of Voice, Vol. ■, No. ■, pp. 1-10
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2015.05.022

The numerous dysphonia self-assessment questionnaires available represent conceptual differences, such as focusing on the handicap,[8] the quality of life,[9] the loss of vocal endurance,[10] or even on the identification of vocal symptoms.[12]

These instruments must be submitted to a series of controlled procedures to be used in a language other than the original. Procedures include linguistic and cultural adaptation, validity, reproducibility, responsiveness, and reliability measures.[14]

Many vocal self-assessment questionnaires that were originally developed in English have been already validated into Brazilian Portuguese by the group of authors of this study. Among these instruments are Qualidade de Vida em Voz (QVV)[15] (Voice-Related Quality of Life [V-RQOL]),[9] Índice de Desvantagem Vocal (IDV)[16] (Voice Handicap Index [VHI; Jacobson et al 1999]),[8] and its reduced version (IDV-10[17] e VHI-10),[18] Questionário de Performance Vocal (QPV)[19] (Vocal Performance Questionnaire [VPQ]),[10] and Escala de Sintomas Vocais (ESV)[20] (Voice Symptom Scale [VoiSS]).[12]

As a result of their validation studies, the mean scores of all the aforementioned questionnaires are known for both dysphonic and vocally healthy individuals.[15,16,19,17,20] These values are presented in this article in Table 1. However, as of now, the literature has not presented sufficient information about the cutoff values of these questionnaires that discriminate healthy from disordered individuals and about the degree of classification based on their efficiency. One of the most adequate analysis used for determining the discriminatory power of a binary classification system, that is, its efficiency in the task it was developed to do, is the receiver operating characteristic (ROC) curve, known as ROC curve. This analysis is a statistical procedure originated from the signal detection theory used for measuring the accuracy of sensorial judgments.[21]

The ROC curve represents the relationship between the sensitivity (ability of a test to correctly identify individuals with the problem in question, ie, the ratio of true positives) and the specificity (ability of a test to correctly identify individuals without the problem in question, ie, the ratio of true negatives) of any given test. The ROC curve is a simple analytical procedure for determining the real value from which two categories are discriminated.[22] This analysis considers the highest possible values of sensitivity and specificity, concomitantly combined with the highest values of efficiency (ability of a test to correctly identify both the positive and negative cases, ie, both the presence and absence of the illness or disorder) and product (a value that confirms the efficiency of the test and must accompany the efficiency's values to indicate greater accuracy of results) producing what is called the cutoff value.

The maximum value of 1.0 for sensitivity, specificity, efficiency, and product indicates that the instrument is able to perform a perfect classification as to what it is proposed to evaluate.

The cutoff value is a number from which the result of a test is classified either as positive (presence of deviation, disorder or illness that is being tested) or negative (absence of what is being tested). If the result found is smaller than the cutoff value, the result of a test is classified as negative and vice versa. The ROC curve allows for the comparison of several diagnostic tests, which is one of its most important applications. To determine whether two ROC curves are equal or different, the area under the curve (AUC) is calculated. This area classifies the level of accuracy of a diagnostic test. Consequently, the AUC measures the performance of the test, for instance in the voice area, its accuracy to identify individuals with voice problems. A test that is not able to discriminate between individuals with or without a certain disorder has an AUC of 0.5 (casual identification). Only areas that have values >0.7 are considered satisfactory.

There are only few studies that determined the ROC curve of self-assessment instruments that evaluate the impact of a dysphonia on the individual's life. All these studies have used specifically the original version of VHI. The studies that used the VHI with 30 items showed cutoff values that vary from 12 to 20. All studies found AUCs that were at least satisfactory. The first study[48] compared patients with glottic cancer and benign laryngeal lesions. They used the Dutch version of the questionnaire and found a cutoff value of 15 points, with sensitivity of 0.97 and specificity of 0.86. Two other research groups analyzed the German[23] and the Polish[24] versions of the VHI in patients with several different types of dysphonia. The cutoff value obtained was 12 points for the Polish study with sensitivity of 0.98 and specificity of 0.95. The Swedish version of the VHI[25] obtained a higher cutoff value of 20 points with a sensitivity of 0.77 and a specificity of 0.87. However, the authors highlighted that one of the limitations of the study

TABLE 1.
Characteristics of the Validation Studies With Subjects Distribution and Mean Total Scores of the Questionnaires, According to Chronologic Completion of the Validation

Questionnaire	Authors and Year of Validation Into Brazilian Portuguese	Dysphonic Individuals					Vocally Healthy Individuals				
		Male	Female	Total	Mean Age	Average Score	Male	Female	Total	Mean Age	Average Score
V-RQOL	Gasparini and Behlau 2009[15]	19	95	114	41.3	65.9	31	89	120	43.0	98.0
VHI	Behlau et al 2011[16]	14	38	52	42.3	48.1	20	44	64	41.1	4.5
VPQ	Paulinelli et al 2012[19]	19	141	160	41.0	27	31	134	165	33.0	19
VHI-10	Costa et al 2013[17]	6	54	60	46.9	18.6	6	44	50	43.4	1.7
VoiSS	Moreti et al 2014[20]	56	104	160	43.0	49.4	49	91	140	42.2	7.1

was the reduced number of participants and that the cutoff values found should not be used solely as a generic reference. For the Persian version of the VHI (in Farsi), the cutoff value was 14.5 with a sensitivity of 92% and a specificity of 95%.[26] There have been only two American studies that looked at the cutoff values for the complete version of the VHI. Interestingly enough, they brought forth different values, probably due to differences in the population studied. One of these studies[27] examined only individuals with organic dysphonia after thyroidectomy and found a cutoff value of 18 points. The other study examined only women[28] and found a cutoff value of 11.5. There is only one study performed with the reduced version of the VHI, the VHI-10, available in the literature.[29] The authors of that study analyzed data from 156 questionnaires and obtained a cutoff value of 11 points. Studies carried out with other questionnaires only provided the mean scores of the population studied but not the cutoff values (Kupfer et al 2012)[30].

Therefore, the purpose of the present study was to determine the cutoff value of five different self-assessment questionnaires that evaluate the impact of a voice problem and to compare their discriminatory power, on the basis of their sensitivity and specificity. The instruments used were V-RQOL, the original and reduced version of the VHI and VHI-10, VPQ, and VoiSS. It is important to highlight two aspects; one is that all these questionnaires were previously validated into Brazilian Portuguese and that the VoiSS' cutoff value was also presented before.[20]

METHODS

This study was approved by the institutional review board of Universidade Federal de Sao Paulo—IRB-UNIFESP #0911/11. All participants signed a consent form.

Original data from validation studies of 975 subjects, "all adults," 486 with different degree and type of vocal disorders (individuals with vocal complaint and dysphonia diagnosed by an otolaryngologist) and 489 vocally healthy individuals (people without vocal complaint, who were either accompanying patients with voice disorders in the hospital or with dermatological/ophthalmological complaints that were seen in the same institutions where the dysphonic patients were selected from. No professional singers were included. Both groups were similar according to age; dysphonic group mean age was 41.9 and vocally healthy individuals was 39.8; ranging from 18 to 81 years. Participants' characteristics are presented in Table 1. Data from the different validation studies were considered as a single database to determine the cutoff values. Table 1 presents the characteristics of the groups. Participants from the validation study of the VHI-10[17] were not included. The responses from the validation study of the original version of the VHI were used for deriving the cutoff values for both the VHI-30 and VHI-10 versions. This was done based on the fact that the data from the VHI study were bigger. Consequently, the total number of questionnaires analyzed was 975, 486 from individuals with dysphonia (378 women and 108 men) and 489 from vocally healthy individuals (358 women and 131 men). Participants' age ranged from 18 to 88 years, with a mean age of 41.9 years for the group with dysphonia and 39.8 years for the vocally healthy group. Demographic characteristics of sex and age were similar for both vocally healthy and dysphonic groups, for each one of the instruments ($P > 0.05$). The higher number of women than men who participated in all the studies reflects the reality of the voice clinic.[31–33]

All questionnaires were answered by the participants after they were examined by the otolaryngologist, or during the voice evaluation, or at the first voice therapy session if voice rehabilitation was indicated. The medical diagnoses were diverse and included cases of both behavioral and organic dysphonia. This aspect was not controlled. Acute voice disorders due to upper airway infections were not included.

Data from the validation studies of the following questionnaires were analyzed: V-RQOL, VHI and VHI-10, VoiSS, and VPQ. The V-RQOL has 10 items and is the first instrument to have been validated into Brazilian Portuguese.[15] The original version of the VHI has 30 items,[16] and the reduced version VHI-10 has 10 items.[17] The VoiSS has 30 questions,[20] and the VPQ has 12 questions.[19] All validations were carried out by the same group of researchers and coordinated by the first author (M.B.) of the present study.

The results of each of the questionnaires were submitted to the ROC curve analysis. The AUC was also calculated for the cutoff values that best discriminated the individuals with and without the disorder. In addition, the best values of sensitivity (which indicates the biggest fraction of true positives in relation to all positives) and specificity (which indicates the biggest fraction of false positives in relation to all negatives) were identified. A point on the upper left hand corner of the ROC curve graph indicates 100% sensitivity and 100% specificity. These values represent a perfect classifier (absence of false positives or negatives), which reflects an ideal instrument. A binary instrument can present the following levels of classification: excellent (0.9–1), good (0.8–0.9), fair (0.7–0.8), and poor (0.6–0.7). Instruments that obtain values between 0.5 and 0.6 fail to be a discriminating system. Because of the fact that the VHI-10 is the most internationally disseminated questionnaire,[18,29,13] its individual items were submitted to an additional ROC curve analysis to find the discriminating power of each question.

RESULTS

The results of the questionnaires' discriminatory power analysis are presented in Table 2 and Figure 1. Two of the questionnaires showed to have a perfect classification power: the VoiSS and the VHI, followed by the VHI-10. The V-RQOL and the VPQ showed less efficiency. The cutoff values for the instruments are (Table 2): VoiSS = 16, VHI = 19, VHI-10 = 7.5, V-RQOL = 91.25, and VPQ = 20.5.

The efficiency values of the questionnaires are presented in Tables 3–7. Tables 3 and 4 summarize the VoiSS and VHI questionnaires' results. They were found to be perfect classifiers. These two instruments also produced a product of 1.0, which confirms their efficiency. The VPQ presented a smaller product (0.685) than its efficiency (0.828).

TABLE 2.
Discriminatory Power of the Self-Assessment Questionnaires—VoiSS, VHI, VHI-10, V-RQOL, and VPQ

Questionnaire	Cutoff Value of the Total Score	Sensitivity	Specificity	Efficiency	Product	Cutoff Value for the Maximum Sensitivity	Cutoff Value for the Maximum Specificity
VoiSS	16	1.000	1.000	1.000	1.000	16	16
VHI	19	1.000	1.000	1.000	1.000	19	19
VHI-10	7.5	0.981	1.000	0.991	0.981	5	7.5
V-RQOL	91.25	0.967	0.860	0.914	0.832	86.25	98.75, maximum specificity 0.947
VPQ	20.5	0.831	0.824	0.828	0.685	15.5	31.5

Notes: ROC curve analysis.

Table 5 summarizes VHI-10 results. The VHI-10 presented its maximum sensitivity at 5.0 points and maximum specificity at 7.5 points. Consequently, if the goal of the screening is to include all individuals with dysphonia using the VHI-10, it is recommended to use the smallest cutoff value (in this case is 5.0 points), which is a more rigorous rating. On the other hand, if the option is to assume the risk of having false negative results, the cutoff value adopted can be 7.5 points.

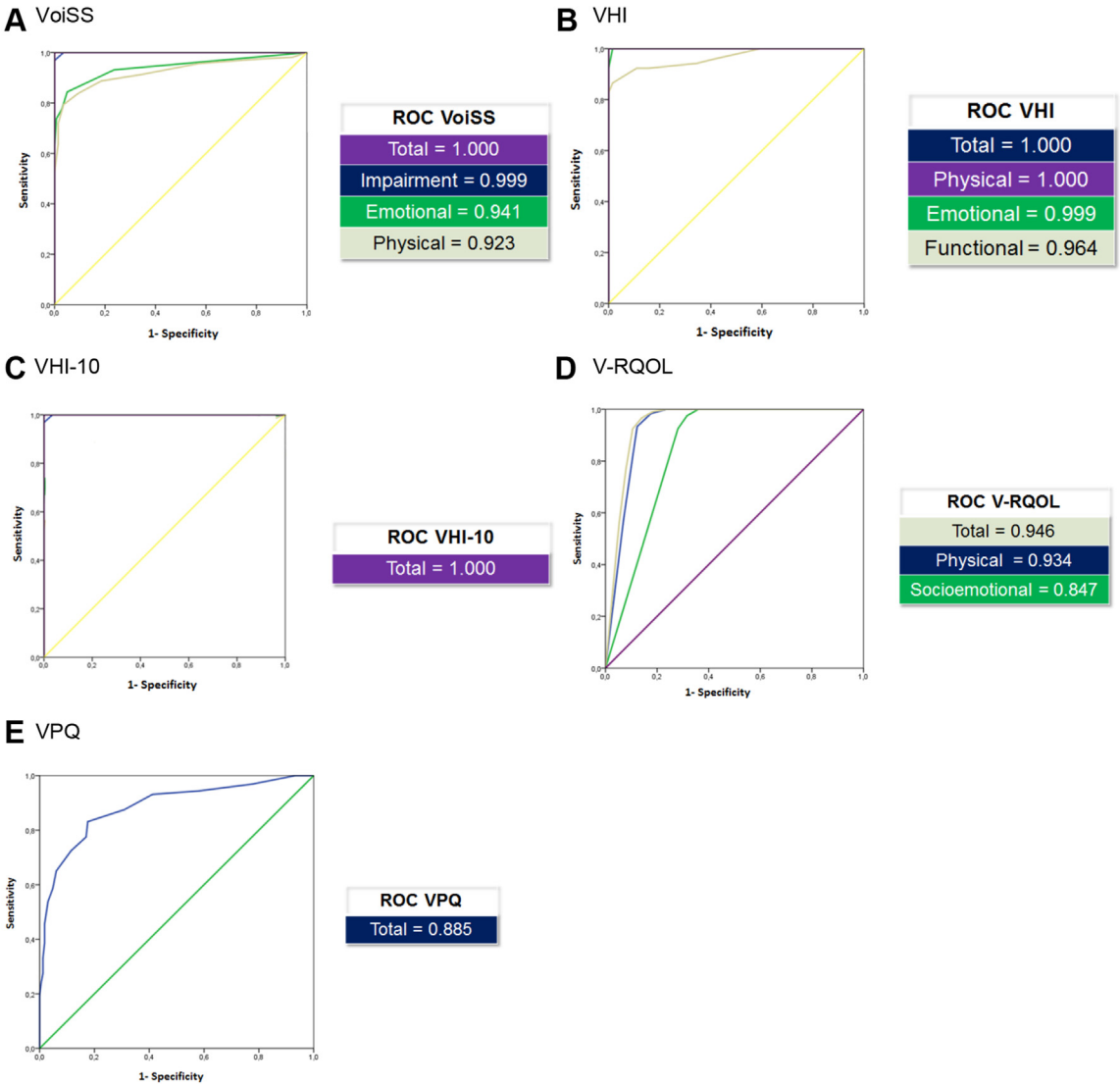

FIGURE 1. ROC curve of the questionnaires VoiSS, VHI, VHI-10, V-RQOL, and VPQ.

TABLE 3.
Cutoff Values, Sensitivity, Specificity, Efficiency, and Product of the VoiSS

Cutoff Value	Sensitivity	1 – Specificity	Specificity	Efficiency	Product
12.500	1.000	0.050	0.950	0.975	0.950
13.500	1.000	0.014	0.986	0.993	0.986
16.000	**1.000**	**0.000**	**1.000**	**1.000**	**1.000**
18.500	0.988	0.000	1.000	0.994	0.988
19.500	0.981	0.000	1.000	0.991	0.981

Bold numbers correspond to the cuttoff values with the highest sensititvity and specificity.

V-RQOL results are presented in Table 6. Results revealed a cutoff value of 91.25 points; this score represented the best relationship between sensitivity and specificity. However, the V-RQOL's cutoff value for maximum sensitivity was 86.25 points. This questionnaire did not produce a maximum specificity cutoff value. Hence, this instrument shows to be better at identifying individuals with voice problem instead of excluding those that do not have it.

Finally, the VPQ findings are presented in Table 7. The results revealed this instrument to be the least efficient at discriminating patients with dysphonia from vocally healthy individuals. A cutoff value of 20.5 represented the point of best relationship between its sensitivity and specificity. The scores for maximum sensitivity and specificity were 15.5 and 31.5, respectively. The latter value is too distant from the value that identifies dysphonic individuals (sensitivity value).

The analysis of the discriminatory power of the individual items of the VHI-10 (Table 8 and Figure 2) showed that 9 of 10 items have discriminatory power. However, item 22 had a poor discriminatory power. In addition, one of the items (item 19) did not present power to discriminate dysphonic patients from vocally healthy individuals.

DISCUSSION

Self-assessment measures that evaluate the impact of a voice problem are a functional approach and key component for voice diagnosis as well as for the therapeutic management of individuals with voice problem. Data provided by these instruments differ from the ones obtained by objective measures and do not necessarily have a direct correlation with perceptual and/or acoustic deviations.[5] The identification of the instrument's efficiency, cutoff value, sensitivity, and specificity is important in order for these instruments to be used as screening tools of large populations, multicentric studies, public service, and/or as a criterion for managing waiting lists and emergency care.[23]

The overall discriminatory power of the instruments validated into Brazilian Portuguese[15,16,19,17,20] was found to be at least good. This fact reinforces the usefulness of these measures for screening big populations as well as for managing decision making. The discriminatory power can assure the reliable use of these measures in the clinical setting because the perceptual analysis has only a poor correlation with the patient's perception of his/her voice problem.[3,4] Some of the instruments produced cutoff values with decimal places; therefore, it is the clinicians' choice as to how rigorous they want to be by choosing either the immediate higher or lower value.

Two questionnaires obtained a perfect classification power, that is, the same cutoff value achieved 100% sensitivity and specificity. These questionnaires were the VoiSS, which has 30 questions and three subscales (maximum efficiency at 16 points) and the VHI, which also has 30 questions and three subscales (maximum efficiency at 19 points).

A detailed analysis of these instruments published at the end of the first decade of the 21st century included nine questionnaires, which met the inclusion criteria established by the authors.[34] The analysis revealed that the quality of these instruments varies greatly, especially in regard to their development process. None of the instruments met the recommended criteria. Flaws in the development process of these instruments have been already identified, more specifically regarding the generation of the initial list of items and the reduction process of the items. In addition to the aforementioned issues, deficiencies were also found in the following aspects: limited breadth of

TABLE 4.
Cutoff Values, Sensitivity, Specificity, Efficiency, and Product of the VHI

Cutoff Value	Sensitivity	1 – Specificity	Specificity	Efficiency	Product
13.000	1.000	0.031	0.969	0.985	0.969
14.500	1.000	0.016	0.984	0.992	0.984
19.000	**1.000**	**0.000**	**1.000**	**1.000**	**1.000**
24.500	0.962	0.000	1.000	0.981	0.962
27.000	0.942	0.000	1.000	0.971	0.942

Bold numbers correspond to the cuttoff values with the highest sensititvity and specificity.

TABLE 5.
Cutoff Values, Sensitivity, Specificity, Efficiency, and Product of the VHI-10

Cutoff Value	Sensitivity	1 − Specificity	Specificity	Efficiency	Product
−1.000	1.000	1.000	0.000	0.500	0.000
1.000	1.000	0.469	0.531	0.766	0.531
2.500	1.000	0.156	0.844	0.922	0.844
3.500	1.000	0.094	0.906	0.953	0.906
5.000	**1.000**	0.031	0.969	0.985	0.969
7.500	**0.981**	**0.000**	**1.000**	**0.991**	**0.981**
9.500	0.962	0.000	**1.000**	0.981	0.962
10.500	0.923	0.000	1.000	0.962	0.923

Bold numbers correspond to the cuttoff values with the highest sensititvity and specificity.

dysphonia types, reduced number of interviews, focus group analysis, and deficits in the psychometric measures. However, the paper in question highlights that the VoiSS is the instrument that has the best structure compared with the others studied. This questionnaire went through a very complex process of construction. It included a larger number of participants (more than 800 subjects). It was considered the most robust measure by the meticulously analytical article by Branski et al.[34] Bearing this in mind, the results of the VHI were surprising considering that it was the first self-assessment instrument developed in the area. It was created before the Scientific Advisory Committee recommendations were published.[14] Both the VoiSS and the VHI produced maximum values of efficiency and product, which confirms their perfect discriminatory ability. The VHI-10 showed to have the third best discriminatory ability. Its efficiency was slightly smaller than the two first instruments although it produced an AUC of 1 (efficiency = 0.991; specificity = 1, sensitivity = 0.981, and cutoff value of 7.5 points). The VHI-10's maximum specificity had a cutoff value of 7.5. This value decreased to five points when considering the maximum sensitivity, which is the cutoff value that should be used when the intention is to identify all individuals with voice problems.

The V-RQOL is the questionnaire that has the most data with Brazilian population.[35] It has 10 items and two subscales. Results of this study showed that the V-RQOL had an excellent AUC (efficiency = 0.914; specificity = 0.860; sensitivity = 0.967; cutoff value = 91.25). Its maximum sensitivity was obtained at a cutoff value of 86.25. However, the V-RQOL did not achieve maximum specificity, which is predicted to be at 101 points (cutoff value obtained was 86.25 points for the highest specificity of 0.947). This means that there is a small chance that a vocally healthy individual will be included in a group of dysphonic patients by using 86.25 as the cutoff value.

Lastly, the VPQ is an instrument that has 12 items and a single total score. In the present study, it showed a good discriminatory power (efficiency = 0.828; specificity = 0.824; sensitivity = 0.831; cutoff value = 20.5 points). The calculation for this instrument's maximum sensitivity and specificity revealed values that are far apart (15.5 and 31.5 points, respectively). This indicates that there is a higher chance for including false positive and negative results with the VPQ during a screening. Hence, this is the least appropriate instrument for a vocal screening. However, it can still be used in the clinical setting, especially when the patient has a clear perception of how his/her voice was like before the onset of the voice problem. The result suggests that this questionnaire may be more useful for the evaluation of organic dysphonia, such as vocal fold paralysis and spasmodic dysphonia. The use of the VPQ should be avoided in cases of long-term behavioral dysphonia in which the patients have difficulty recollecting how their voice was before the onset of the problem.

The analysis of the relationship between the number of items of the questionnaire (length of the instrument) and the time of administration and its efficiency indicated that the VHI-10 is the best option. The study of the individual items' efficiency

TABLE 6.
Cutoff Values, Sensitivity, Specificity, Efficiency, and Product of the V-RQOL

Cutoff Value	Sensitivity	1 − Specificity	Specificity	Efficiency	Product
83.750	1.000	0.263	0.737	0.869	0.737
86.250	**1.000**	0.237	0.763	0.882	0.763
88.750	0.992	0.184	0.816	0.904	0.809
91.250	**0.967**	**0.140**	**0.860**	**0.914**	**0.832**
93.750	0.925	0.105	0.895	0.910	0.828
96.250	0.767	0.079	0.921	0.844	0.706
98.750	0.558	0.053	0.947	0.753	0.528
101.000	0.000	0.000	**1.000**	0.500	0.000

Bold numbers correspond to the cuttoff values with the highest sensititvity and specificity.

TABLE 7.
Cutoff Values, Sensitivity, Specificity, Efficiency, and Product of the VPQ

Cutoff Value	Sensitivity	1 − Specificity	Specificity	Efficiency	Product
14.500	1.000	0.952	0.048	0.524	0.048
15.500	**1.000**	0.933	0.067	0.534	0.067
16.500	0.969	0.776	0.224	0.597	0.217
17.500	0.944	0.582	0.418	0.681	0.395
18.500	0.931	0.412	0.588	0.760	0.547
19.500	0.875	0.309	0.691	0.783	0.605
20.500	**0.831**	**0.176**	**0.824**	**0.828**	**0.685**
21.500	0.775	0.170	0.830	0.803	0.643
22.500	0.725	0.115	0.885	0.805	0.642
23.500	0.650	0.061	0.939	0.795	0.610
24.500	0.588	0.048	0.952	0.770	0.560
25.500	0.538	0.030	0.970	0.754	0.522
26.500	0.456	0.018	0.982	0.719	0.448
27.500	0.388	0.018	0.982	0.685	0.381
28.500	0.331	0.012	0.988	0.660	0.327
29.500	0.275	0.012	0.988	0.632	0.272
30.500	0.244	0.006	0.994	0.619	0.243
31.500	0.194	0.000	**1.000**	0.597	0.194
32.500	0.150	0.000	1.000	0.575	0.150

Bold numbers correspond to the cuttoff values with the highest sensititvity and specificity.

of the VHI-10 showed that an item failed to discriminate dysphonic patients from vocally healthy individuals ("I feel left out of conversations because of my voice"), with an unsatisfactory AUC (=0.538). This low value reflects that the item is not able to perform casual identification. Perhaps, this unsatisfactory value is due to cultural aspects related to communication or to the fact that this item is more sensitive to disorders in which there is a marked voice deviation, such as neurologic dysphonia or head and neck cancer. In addition, there was another item that had a poor discrimination ("my voice problem causes me to lose income"), with an area under the curve considered just satisfactory (AUC = 0.702). Probably, the low value of this item is due to the fact that it may be more sensitive to professional voice users. The participants of this study were mostly from hospitals and university clinics. All remaining items had an AUC higher than 0.8. One of these items was found to be a perfect classifier with an AUC of 1.0 ("my voice problem upsets me"). Although the purpose of this analysis was to understand the individual contribution of

TABLE 8.
Area Under the Curve (AUC) for the VHI Items That Compose the VHI-10 and Total Score

Question	AUC	Standard Error	*P* Value
Q_01	0.883	0.035	<0.001*
Q_03	0.896	0.029	<0.001*
Q_10	0.986	0.012	<0.001*
Q_14	0.922	0.029	<0.001*
Q_16	0.875	0.037	<0.001*
Q_17	0.814	0.044	<0.001*
Q_19	0.538	0.054	0.477
Q_22	0.702	0.051	<0.001*
Q_23	1.000	0.000	<0.001*
Q_25	0.827	0.043	<0.001*
Total	1.000	0.001	<0.001*

*Significant values—ROC curve analysis.
Item: 1. My voice makes it difficult for people to hear me.
Item: 3. People have difficulty understanding me in a noisy room.
Item: 10. People ask, "What's wrong with your voice?"
Item: 14. I feel as though I have to strain to produce voice.
Item: 16. My voice difficulties restrict personal and social life.
Item: 17. The clarity of my voice is unpredictable.
Item: 19. I feel left out of conversations because of my voice.
Item: 22. My voice problem causes me to lose income.
Item: 23. My voice problem upsets me.
Item: 25. My voice makes me feel handicapped.

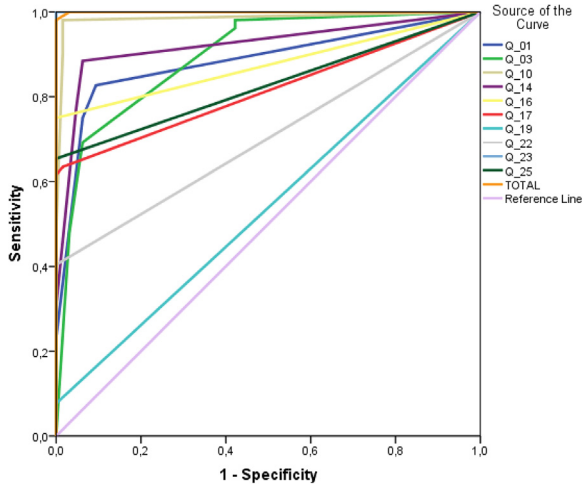

FIGURE 2. ROC curve for the VHI items that compose the VHI-10 and total score.

each item of the instrument, it can suggest that a reduction of the numbers of items may be appropriate so that an efficient shorter version of this questionnaire can be available.

Ultimately, it is the clinician's decision to choose the cutoff values of interest, given that during screenings, it would be more important to consider the sensitivity values that are smaller or equal to the efficiency values of the instrument. This way, the highest number of true positives in relation to all positives is included. This is not a problem with the instruments that achieved either maximum efficiency or sensitivity, such as the VoiSS and the VHI. However, when the selected instruments are the VHI-10, the V-RQOL, and the VPQ, either the maximum efficiency or sensitivity or even specificity values can be selected. In the clinical setting, the value that should be considered is the maximum efficiency because there are other data available in the voice evaluation to support the diagnosis. In the case of a population screening, the choice will vary according to the amount of sensitive or specific of the goal of the screening itself.

The VoiSS and the VHI presented a single cutoff value. The VHI-10, V-RQOL, and VPQ have variable thresholds depending on the purpose of the screening. The knowledge of the efficiency of assessment instruments and treatment protocols can help optimize speech-language pathology intervention during screenings and clinical management.

After the advent of objective voice analysis, the era of voice self-assessment instruments "represents progress" in the evaluation of voice patients, despite the fact that there are flaws in their development.[36,34]

The establishment of cutoff values of several validated questionnaires into Brazilian Portuguese may improve the clinical ability to detect the presence or progress of a disease or disorder as well as to measure treatment outcome.

The positive result of the discriminatory power of these instruments—V-RQOL, VHI, VHI-10, VPQ, and VoiSS—allows for the ratification of their diagnostic usefulness. Moreover, their administration does not depend on the level and experience of the clinician, which are fundamental criteria for perceptual analysis[37,38] (Kreiman et al 2007). Self-assessment questionnaires can also be ideal tools for international multicentric researches, as they do not depend on the education and experience of the clinician nor on the influence of his/her native language, contrary to perceptual analysis.[37] There are limited studies that investigate the role of culture on the perception of a dysphonia.[39] Recently, a group from India proposed a culturally developed questionnaire called the voice disorder outcome profile.[40] The study considered environmental and cultural aspects of India, such as spicy food, excessive tea and coffee intake, tropical weather, excessive voice use in the streets and over background noise, pollution, and lack of acoustic amplification. This questionnaire was translated and validated into Tamil.[41] Aside from the specific cultural items, the remaining items are very similar to those in other self-assessment questionnaires. The correlation between a universally established self-assessment tool such as the VHI, and instruments that include particular aspects, such as culture, is not clear at this juncture and deserves further investigation.

The VHI is undoubtedly the questionnaire that has been mostly validated in other languages other than the original.[42,13] However, other instruments have been successfully culturally adapted and validated into other languages as well, such as the V-RQOL[43,15] and the VAPP[44–46] (Fava et al 2014).

The scores of the questionnaires and the analysis of their cutoff values may allow for the evaluation of the biosocial impact of a dysphonia and support their use as screening tools. Despite the fact that we did not include several diagnostic categories in the validation studies of these questionnaires, previous studies proved that one of the limitations imposed by a voice problem is that patients perceive their voice the same way regardless of the disorder itself (eg, patient with head and neck cancer or a functional disorder).[47]

Professional voice users may self-rate their voice problem differently. They may be more detailed or use a more restrict criteria, even in cases of very mild loss of vocal functioning.[49] It is known that self-assessment instruments that are not specific to professional voice users are less sensitive in evaluating this population.[50,49,51] However, we did not control for profession as a variable in our studies and the 1000 plus participants had several different professions, some of whom were not professional voice users.

We understand that none of these instruments were specifically developed for the evaluation of a peculiar diagnostic category; however, they reflect a subjective perception of the limitations or handicap imposed by a voice problem. We highlight once more that although there is a correlation between the self-assessment and the perceptual analysis, this correlation is not direct nor strong.[8,52,53,54,25,4] Moreover, a huge vocal deviation perceived by a nonprofessional voice user may be the same as a mild voice deviation for ones that rely on their voice to work. It is possible that the self-assessment instrument captures a different aspect of the vocal function that the perceptual analysis and laryngeal examination do not. Therefore, it is recommended that clinicians use all information consciously during the evaluation process and the treatment outcome measure and discharge because they are complementary tools.

CONCLUSIONS

The results of the present study indicate that self-assessment questionnaires are useful clinical instruments, which are valid and reliable to differentiate individuals with voice problems from vocally healthy individuals. The VoiSS and VHI are perfect classifiers because they obtained maximum efficiency. The VHI-10 and V-RQOL are excellent classifiers, and finally the VPQ is a good classifier. Analysis of the VHI-10 revealed the discriminatory power on all but two items for Brazilian speakers—feeling left of out conversations and loss of income because of their voice problem.

Acknowledgment

This work was supported by a grant from the "Conselho Nacional de Desenvolvimento Científico e Tecnológico" - CNPq

(National Counsel fo technological and Scientific Development). The authors would like to express appreccciation.

REFERENCES

1. Schwartz SR, Cohen SM, Dailey SH, et al. Clinical practice guideline: hoarseness (dysphonia). *Otolaryngol Head Neck Surg*. 2009;141(suppl 2):S1–S31.
2. Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol*. 2001;258:77–82.
3. Karnell MP, Melton SD, Childes JM, Coleman TC, Dailey AS, Hoffman HT. Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *J Voice*. 2007;21:576–590.
4. Ugulino AC, Oliveira G, Behlau M. Perceived dysphonia by the clinician's and patient's viewpoint. *J Soc Bras Fonoaudiol*. 2012;24:113–118.
5. Hsiung MW, Pai L, Wang HW. Correlation between voice handicap index and voice laboratory measurements in dysphonic patients. *Eur Arch Otorhinolaryngol*. 2002;259:97–99.
6. Woisard V, Bodin S, Yardeni E, Puech M. The voice handicap index: correlation between subjective patient response and quantitative assessment of voice. *J Voice*. 2007;21:623–631.
7. World Health Organization. *The International Classification of Functioning*. Geneva: Disability and Health; 2001.
8. Jacobson BH, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger MS, Newman CW. The Voice Handicap Index (VHI): development and validation. *Am J Speech Lang Pathol*. 1997;6:66–70.
9. Hogikyan ND, Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *J Voice*. 1999;13:557–569.
10. Carding PN, Horsley IA, Docherty GJ. A study of the effectiveness of voice therapy in the treatment of 45 patients with nonorganic dysphonia. *J Voice*. 1999;13:72–104.
11. Ma EP, Yiu EM. Voice activity and participation profile: assessing the impact of voice disorders on daily activities. *J Speech Lang Hear Res*. 2001;44:511–524.
12. Deary IJ, Wilson JA, Carding PN, MacKenzie K. VoiSS: a patient-derived Voice Symptom Scale. *J Psychosom Res*. 2003;54:483–489.
13. Behlau M, Murry T. International and intercultural aspects of voice and voice disorder. In: Battle D, ed. *Communication Disorders in Multicultural and International Populations*. St Louis, MO: Elsevier; 2012:174–207.
14. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, Stein RE. Scientific Advisory Committee of Medical Outcomes Trust. Assessing health status and quality of life instruments: attributes and review criteria. *Qual Life Res*. 2002;11:193–205.
15. Gasparini G, Behlau M. Quality of life: validation of the Brazilian version of the voice-related quality-of-life (V-RQOL) measure. *J Voice*. 2009;23:76–81.
16. Behlau M, Alves Dos Santos L de M, Oliveira G. Cross-cultural adaptation and validation of the voice handicap index into Brazilian Portuguese. *J Voice*. 2011;25:354–359.
17. Costa T, Oliveira G, Behlau M. Validation of the Voice Handicap Index: 10 (VHI-10) to the Brazilian Portuguese. *Codas*. 2013;25:482–485.
18. Rosen CA, Lee AS, Osborne J, Zullo T, Murry T. Development and validation of the Voice Handicap Index-10. *Laryngoscope*. 2004;114:1549–1556.
19. Paulinelli B, Gama AC, Behlau M. Validation of the vocal performance questionnaire in Brazil. *Rev Soc Bras Fonoaudiol*. 2012;17:85–91.
20. Moreti F, Zambon F, Oliveira G, Behlau M. Cross-cultural adaptation, validation, and cutoff values of the Brazilian version of the Voice Symptom Scale-VoiSS. *J Voice*. 2014;28:458–468.
21. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006; 27:861–874.
22. Carter RE, Lubinsky J, Domholdt E. *Rehabilitation Research: Principles and Applications*. New York: Elsevier Saunders; 2000.
23. Gräßel E, Hoppe U, Rosanowski F. Grading of the Voice Handicap Index. *HNO*. 2008;56:1221–1228.
24. Niebudek-Bogusz E, Kuzańska A, Woznicka E, Sliwinska-Kowalska M. Assessment of the voice handicap index as a screening tool in dysphonic patients. *Folia Phoniatr Logop*. 2011;63:269–272.
25. Ohlsson AC, Dotevall H. Voice handicap index in Swedish. *Logoped Phoniatr Vocol*. 2009;34:60–66.
26. Moradi N, Pourshahbaz A, Majid S, Javadipour S. Cutoff point at voice handicap index used to screen voice disorders among Persian speakers. *J Voice*. 2012;27:130.e1–130.e5.
27. Solomon NP, Helou LB, Henry LR, Howard RS, Coppit G, Shaha AR, Stojadinovic A. Utility of the Voice Handicap Index as an indicator of postthyroidectomy voice dysfunction. *J Voice*. 2013;27:348–354.
28. Behrman A, Rutledge J, Hembree A, Sheridan S. Vocal hygiene education, voice production therapy, and the role of patient adherence: a treatment effectiveness study in women with phonotrauma. *J Speech Lang Hear Res*. 2008;51:350–366.
29. Arffa R, Krishna P, Gartner-Schmidt J, Rosen CA. Normative values for the voice handicap index-10. *J Voice*. 2012;26:462–465.
30. Kupfer R, Hogikyan E, Hogikyan N. Establishment of a normative database for the Voice-Related Quality of Life (V-RQOL) measure. *J Voice*. 2013;28: 449–451.
31. Titze IR, Lemke J, Montequin D. Populations in the U.S. workforce who rely on voice as a primary tool of trade: a preliminary report. *J Voice*. 1997;11:254–259.
32. Roy N, Merrill RM, Gray SD, Smith EM. Voice disorders in the general population: prevalence, risk factors, and occupational impact. *Laryngoscope*. 2005;115:1988–1995.
33. Cohen SM, Kim J, Roy N, Asche C, Courey M. Prevalence and causes of dysphonia in a large treatment-seeking population. *Laryngoscope*. 2012; 122:343–348.
34. Branski RC, Cukier-Blaj S, Pusic A, et al. Measuring quality of life in dysphonic patients: a systematic review of content development in patient-reported outcomes measures. *J Voice*. 2010;24:193–198.
35. Behlau M, Hogikyan ND, Gasparini G. Quality of life and voice: study of a Brazilian population using the voice-related quality of life measure. *Folia Phoniatr Logop*. 2007;59:286–296.
36. Franic DM, Bramlett RE, Bothe AC. Psychometric evaluation of disease specific quality of life instruments in voice disorders. *J Voice*. 2005;19: 300–315.
37. Kreiman J, Gerratt BR, Khan SU. Effects of native language on perception of voice quality. *J Phon*. 2010;38:588–593.
38. Sofranko JL, Prosek RA. The effect of levels and types of experience on judgment of synthesized voice quality. *J Voice*. 2014;28:24–35.
39. Yiu EM, Ho EM, Ma EP, Verdolini Abbott K, Branski R, Richardson K, Li NY. Possible cross-cultural differences in the perception of impact of voice disorders. *J Voice*. 2011;25:348–353.
40. Konnai RM, Jayaram M, Scherer RC. Development and validation of a voice disorder outcome profile for an Indian population. *J Voice*. 2010; 24:206–220.
41. Mahalingam S, Boominathan P, Subramaniyan B. Voice Disorder Outcome Profile (V-DOP)—translation and validation in Tamil language. *J Voice*. 2014;28:841.e21–841.e32.
42. Verdonck-de Leeuw IM, Kuik DJ, De Bodt M, et al. Validation of the voice handicap index by assessing equivalence of European translations. *Folia Phoniatr Logop*. 2008;60:173–178.
43. Schwanfelder C, Eysholdt U, Rosanowski F, Graessel E. Voice-related quality of life: structure, validity and factors of the German questionnaire. *Folia Phoniatr Logop*. 2008;60:241–248.
44. Kleemola L, Helminen M, Rorarius E, Isotalo E, Sihvo M. Voice activity and participation profile in assessing the effects of voice disorders on quality of life: estimation of the validity, reliability and responsiveness of the Finnish version. *Folia Phoniatr Logop*. 2011; 63:113–121.
45. Ricarte A, Oliveira G, Behlau M. Validation of the Voice Activity and Participation Profile protocol in Brazil. *Codas*. 2013;25:242–249.
46. Fava G, Paolillo NP, Oliveira G, Behlau M. Cross-cultural adaptation, validation, and cutoff point of the Italian version of the voice activity and participation profile: Profilo di Attività e Partecipazione Vocale. *J Voice*. 2015;29: 130.e11–130.e19.

47. Weigelt S, Krischke S, Klotz M, Hoppe U, Köllner V, Eysholdt U, Rosanowski F. Voice handicap in patients with organic and functional dysphonia. *HNO*. 2004;52:751–756.
48. Van Gogh CD, Mahieu HF, Kuik DJ, Rinkel RN, Langendijk JA, Verdonck-de Leeuw IM. Voice in early glottic cancer compared to benign voice pathology. *Eur Arch Otorhinolaryngol*. 2007;264:1033–1038.
49. Murry T, Zschommler A, Prokop J. Voice handicap in singers. *J Voice*. 2009;23:376–379.
50. Rosen CA, Murry T. Voice handicap index in singers. *J Voice*. 2000;14:370–377.
51. Paoliello K, Oliveira G, Behlau M. Singing voice handicap mapped by different self-assessment instruments. *Codas*. 2013;25:463–468.
52. Nawka T, Wiesmann U, Gonnermann U. Validation of the German version of the voice handicap index. *HNO*. 2003;51:921–930.
53. Speyer R, Wieneck GH, Dejonckere PH. Self-assessment of voice therapy for chronic dysphonia. *Clin Otolaryngol Allied Sci*. 2004;29:66–74.
54. Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen IN, Wilson JA. Optimising outcome assessment of voice interventions, I: reliability and validity of three self-reported scales. *J Laryngol Otol*. 2007;121:763–767.