The German Restructured Vocal Fatigue Index and Characteristics of Dysphonic and Vocally Healthy Populations

[*,†]Ben Barsties v. Latoszek, [‡]Margrit Göllner, [†,a]Philipp Mathmann, and [†,a]Katrin Neumann, *Düsseldorf, †Münster, and ‡Nümbrecht, Germany

Summary: Objective. The Vocal Fatigue Index (VFI), a 19-item psychometric self-report questionnaire, enables individuals with vocal fatigue (VF) to be identified and their complaints to be characterized. The purpose of this study was to improve the German-language version (VFI-G) and to evaluate further vocal fatigue-related characteristics of dysphonic and control populations.

Methods. The VFI-G was restructured by replacing the three factors that structured the original: (1) tiredness of voice and voice avoidance; (2) physical discomfort; and (3) improvement of symptoms with rest, with two clusters developed on the basis of the results of a factor analysis by Nanjundeswaran et al. (2019). The two new clusters are: (1) tiredness and avoidance plus physical discomfort; and (2) symptom improvement through rest. One hundred one (101) individuals with voice disorders and 100 vocally healthy controls from a previous study that cross-validated the VFI-G participated in this study. In order to assess the validity of our newly adjusted VFI-G, independent samples *t* test, receiver operating characteristic curve, likelihood ratios and the Youden Index were calculated. The association of the two VF clusters with subject characteristics such as age, sex, type of voice disorder, and level of vocal usage was also analyzed using either a Pearson correlation or a one-way ANOVA for each of the two populations.

Results. Significantly higher scores were obtained in voice-disordered subjects in both clusters (all P values < 0.001) than in healthy-voice subjects. The threshold for cluster 1 of the VFI-G was determined as ≥17.5 (74.3% sensitivity and 88.0% specificity). The results of cluster 2 are identical to that of factor 3 of the previous cross-validation study of the VFI-G. Most subject characteristics show no significant association with cluster 1 of the VFI-G, but cluster 2 seems to be moderately associated with age, type of voice disorder and level of vocal usage in the dysphonic population.

Conclusions. The restructured VFI-G showed improved validity and can be recommended for use in the assessment of VF. Cluster 2 is also moderately associated with several vocal fatigue-related subject characteristics of the dysphonic population.

Key Words: Vocal fatigue−Self-perception−Questionnaire−Voice−Dysphonia.

INTRODUCTION

Voice problems present a substantial burden and have a significant impact on the quality of life of those affected, especially regarding limitations in their work and social life. Prevalence reports of dysphonia in the general population range from 7% to 17%.[1−3] Members of some professional groups suffer from dysphonia significantly more than others, including occupational voice users, such as teachers, clergy, salespeople, courtroom attorneys, telemarketers and receptionists, and professional voice users, such as singers, actors and broadcast personalities.

The prevalence of voice disorders in teachers, especially kindergarten and primary school teachers, have been reported to range from 27% to 80% over the course of their professional lives.[4−7] Voice problems have been reported to be approximately two to three times more common among teachers than nonteachers, and women in general were reported to have a higher lifetime prevalence of voice disorders than men in general.[3,4,7] An Italian study revealed that 23% of teachers were absent from work due to voice problems but that only 5.4% of a control group who did not have particular demands on their voices were absent for that same reason.[4] Forty-eight to 86% of priests reported voice problems during their careers and 16% often suffered from voice problems.[8,9] One systematic review showed that 46% of singers have voice problems during their career.[10] Voice disorders account for 20% of occupational disease in Poland and are the most frequently-certified category of occupational disease there.[11]

Vocal fatigue is a common symptom in cases of dysphonia. Professional or elite voice users, such as teachers, actors and singers, are especially prone to developing vocal fatigue. Teachers are three times more likely to be affected by vocal tiredness, vocal avoidance and voice discomfort than vocally healthy individuals, in which female teachers being more severely affected than males.[12,13] Understanding the underlying physiological and biomechanical mechanisms of vocal fatigue is therefore important for appropriate

Accepted for publication January 26, 2021.
Conflict of interest: This manuscript has no actual or potential conflict of interest.
From the *Speech-Language Pathology, SRH University of Applied Health Sciences, Düsseldorf, Germany; †Department of Phoniatrics and Pediatric Audiology, University Hospital Münster, University of Münster, Münster, Germany; and the ‡Department of Speech-Language Pathology, Dr. Becker Rhein-Sieg-Klinik, Nümbrecht, Germany.
[a]These authors are co-senior authors on this work.
Address correspondence and reprint requests to Ben Barsties v. Latoszek, Graf-Adolf-Straße 67, 40210 Düsseldorf, Germany. E-mail: benjamin.barstiesvonlatoszek@srh.de
Journal of Voice, Vol. 37, No. 3, pp. 469.e11−469.e18
0892-1997
© 2021 The Authors. Published by Elsevier Inc. on behalf of The Voice Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
https://doi.org/10.1016/j.jvoice.2021.01.024

diagnosis and treatment. A number of mechanisms are believed to be involved in the development of vocal fatigue, including neuromuscular fatigue, increased vocal fold viscosity, reduced blood circulation, nonmuscular tissue strain and respiratory muscle fatigue.[14] But because of the multi-faceted character and unique nature of voice production in human functions, based, as it is, on the repeated acceleration and deceleration of tissue over time, vocal fatigue is not fully understood.[15] This may be one reason why there is no common definition of vocal fatigue.[16,17] The authors of one recent comprehensive literature review tried to find a consensus for this complex phenomenon,[17] claiming that vocal fatigue could be understood as an individually perceived, measurable symptom affecting vocal task performance and integrating many different self-perceived vocal symptoms or physiological deficits, resulting in a probably high-vocal demand response, high vocal effort or neuromuscular deficit.[17]

The diagnosis of vocal fatigue seems to be based mainly on self-evaluation by the patient. Quantification of self-assessed handicap through standardized questionnaires is therefore highly recommended as part of a battery of voice examinations.[18−20] Such standardized questionnaires are particularly suitable for quantifying the extent of activity and participation limitations caused by the effects of vocal fatigue in daily life.

The Vocal Fatigue Index (VFI) is a questionnaire that was developed to identify people with vocal fatigue[21] and has recently been validated or cross-validated in nine different languages.[21−29] The questionnaire has 19 items and measures individual vocal fatigue by analyzing three factors: (1) tiredness of voice and voice avoidance; (2) physical discomfort associated with voicing; and (3) improvement of symptoms with rest.[21]

Factors 1 and 3 are good predictors of the presence or absence of vocal fatigue in people with dysphonia but the value of factor 2 is disputed.[21] Nanjundeswaran et al[30] conducted a recent study to investigate the role of factor 2 in chronic vocal fatigue.[30] In order to determine whether this subscale was conceptually distinct from factor 1, the authors restructured the VFI, seeking to understand the hierarchy of the statements of vocal fatigue.[30] They applied a Mokken scaling, in which the structures of factors 1 and 2 were reorganized into a single cluster (cluster 1).[30] Cluster 2 was being formed by the unchanged factor 3.[30] Cluster 1 and 2 are good predictors of the presence/absence of vocal fatigue, because first, certain functional and physical symptoms of vocal fatigue are not equal factors, which received more robustness, and second, questions of improvement of vocal fatigue were considered with high reliability within items.[30] Mokken scaling assists in reorganizing the motivational structure of communication and providing insight to the complex nature within the underlying trait of fatigue. The restructured VFI requires further language-specific validation and the proposal of threshold values for clinical decision making.

The aim of the present study was to restructure the German-language version of the Vocal Fatigue Index (VFI-G) according to the recommendations by Nanjundeswaran et al.[30] For this purpose, the vocal characteristics of a group of individuals with dysphonia and a vocally healthy control group were evaluated. Factors influencing vocal fatigue, such as age, sex, level of vocal usage, and type of voice disorder, were assessed as co-variables.

METHOD

Subjects

In this study the VFI-G scores of subjects from a previous study by Barsties v. Latoszek et al. (2021)[24] were reanalyzed. The dysphonia group consisted of 101 patients with voice disorders. The participants with voice disorders in this group are more representative of genuine clinical presentations in terms of type and severity distribution of dysphonia, age, and gender distribution than other studies.[31]

The control group consisted of 100 volunteers with no reported history or current symptoms of voice, speech, or hearing problems.

Table 1 summarizes the demographic data, the types of dysphonia and levels of vocal usage (ranging from elite vocal performer to nonvocal nonprofessional) for the dysphonia and control groups.

For further details regarding data acquisition, inclusion and exclusion criteria, test-retest reliability, and previous cross-validation results of the VFI, we refer to Barsties v. Latoszek et al.[24]

Adaption of the VFI-G

The original VFI-G utilized the following three factors in order to assess vocal fatigue[24]: (1) tiredness of voice and voice avoidance; (2) physical discomfort associated with voicing; and (3) improvement of symptoms with rest. Factor 2 and factor 3 revealed the lowest validity levels of this questionnaire. Factor 2 of the VFI-G also showed a significantly lower clinical threshold for discriminating the presence or absence of physical discomfort than other validation studies of the VFI, a finding which may indicate bias.[21,23,28] We, therefore, followed the recommendation of Nanjundeswaran et al[30] to restructure and reorder the VFI-G for further analysis. This adapted version of the VFI contains two clusters: cluster 1 represents the combined factors 1 and 2, and cluster 2 is the unchanged factor 3. The restructured VFI-G is depicted in the Appendix.

Statistics

An independent-sample *t* test was conducted initially between the dysphonia and control groups to verify whether there was acceptable diagnostic discrimination between the two clusters of the restructured VFI-G.

A receiver operating characteristic (ROC) curve was then generated in order to analyze the diagnostic accuracy of the restructured VFI-G according to sensitivity (the scores of

TABLE 1.
Demographic Data, Types of Voice Disorders and Levels of Vocal Usage of the Dysphonia and Control Groups

Group	Type of Dysphonia	Number	Gender		Age in Years Mean ± SD	Level of Vocal Usage
			Female	Male		
Dysphonia Group	Functional dysphonia	55	46	9	46.02 ± 15.66	Level 1: 4%
	Laryngeal neurological dysphonia (i.e., n. vagus paralysis, n. laryngeus recurrens paralysis, and spasmodic dysphonia)	16	7	9	58.63 ± 11.21	Level 2: 15% Level 3: 17% Level 4: 31%
	Inflammation, epithelial lesion	16	11	5	54.69 ± 16.30	Unknown: 33%
	Neoplasm	7	3	4	62.71 ± 8.30	
	Paradoxical vocal fold movement	3	1	2	59.33 ± 12.90	
	External laryngeal trauma	2	0	2	36.50 ± 6.36	
	Dysphonia after radiotherapy	1	1	0	72.00	
	Psychogenic dysphonia	1	0	1	51.00	
Control Group	None	100	71	29	40.78 ± 16.8	Level 1: 15% Level 2: 17% Level 3: 32% Level 4: 36%

the subjects with voice disorders) and specificity (the scores of subjects without voice disorders). The power of the restructured VFI-G to discriminate between subjects with the absence and presence of voice disorders was estimated using the area under the ROC curve (A_{ROC}). An A_{ROC} of >0.9 is considered to be exceptionally good, an A_{ROC} of <0.7 is considered to be low and an A_{ROC} = 0.5 corresponds with a chance-level of diagnostic accuracy.[32] In order to find the optimal threshold value that best differentiates between subjects without and with voice disorders, the Youden index (a measure that uses a receiver operating characteristic to determine which threshold value is best suited to distinguish two groups in a measurement) was calculated as sensitivity + specificity − 1. The likelihood ratio (LR) was calculated in order to verify the applicability of the best threshold for a clinical decision. The LR indicates how many times more or less likely a test result is to occur in individuals who are ill than in those who are healthy, and represents the relationship between sensitivity and specificity. A positive likelihood ratio (LR+) is used for a positive test result and is calculated as the probability of a positive test result in an ill person divided by the probability of a positive test result in a healthy person (LR+ = sensitivity / (1− specificity)); a negative likelihood ratio (L−) is used for a negative test result and is calculated as the probability of a negative test result in an ill person divided by that of a negative test result in a healthy person (LR− = (1−sensitivity)/specificity). By convention LR+ values >10 and LR− values <0.1 indicate large and often conclusive diagnostic evidence, an LR+ between 5 and 10 and an LR− between 0.1 and 0.2 suggest moderate diagnostic evidence, an LR+ between 2 and 5 and an LR− between 0.2 and 0.5 point to a low level of diagnostic evidence, and an LR+ between 1 and 2 and an LR− between 0.5 and 1, respectively, rarely represent relevant diagnostic evidence.[33]

Associations between the type of voice disorder, level of vocal usage and sex of the individuals with the two clusters of the VFI-G were investigated. One-way ANOVAs were run to test the association between specific categorical characteristics (sex, type of voice disorder, level of vocal usage) and the two clusters. These one-way ANOVAs were calculated separately for the dysphonia group and the control group. Levene statistics were used: the Brown-Forsythe where heterogeneity was found, and a regular one-way ANOVA where not. Where significant differences were verified, a post-hoc test was applied to assess pairwise group differences. In this case, we used the Hochberg GT2 because of different sample sizes in the various factors.[34]

The association of age with the two clusters was investigated by calculating the Pearson product-moment correlation within the dysphonia and control groups.

Statistical analyses were performed using SPSS, version 23 for Windows (IBM Corp., Armonk, New York).

RESULTS

Validity

*t*Tests demonstrated that the dysphonic group had significantly higher scores in the two clusters of the VFI-G than the control group (all P-values < 0.001) (Table 2).

Table 3 shows the optimal thresholds for clusters 1 and 2, calculated on the basis of the Youden index and their diagnostic strengths, such as sensitivity, specificity and LR± statistics, all of which were derived from the ROC curve (Figure 1). Cluster 1 of the VFI-G demonstrates an acceptable precision for differentiating between subjects without and with voice disorders (A_{ROC} = 0.846). The threshold of ≥17.5 showed a moderate sensitivity of 74.3% and a high specificity of 88.0%, and the LR values indicated moderate diagnostic evidence. The results of cluster 2 are identical to

TABLE 2.
Clustered VFI-G Values in the Dysphonic and Control Groups

Group	Cluster 1		Cluster 2	
	Mean	SD	Mean	SD
Control Group	10.29	7.98	4.50	3.42
Dysphonia Group	25.78	13.19	6.65	3.49

those of factor 3 from the previous cross-validation study, in which the specificity of 80% was acceptable but the sensitivity of 50.5% was unsatisfactory.[24]

VFI-G clusters related to age

The association between age and each of the two clusters was negligible in both groups (Table 4).

VFI-G clusters related to sex

The homogeneity of VFI-G variances related to sex was asserted using Levene's Test, which showed that heterogeneous variances had to be assumed for cluster 1 ($P = 0.027$) and equal variances for cluster 2 ($P= 0.470$) in the dysphonia group. The significance from the ANOVA was therefore used for cluster 2 and from the Brown-Forsythe for cluster 1. Sex had no significant impact on the results of either cluster 1 ($P = 0.523$) or cluster 2 ($P = 0.070$).

The Levene's test of the control group data did not indicate heterogeneity of VFI-G variances related to sex for

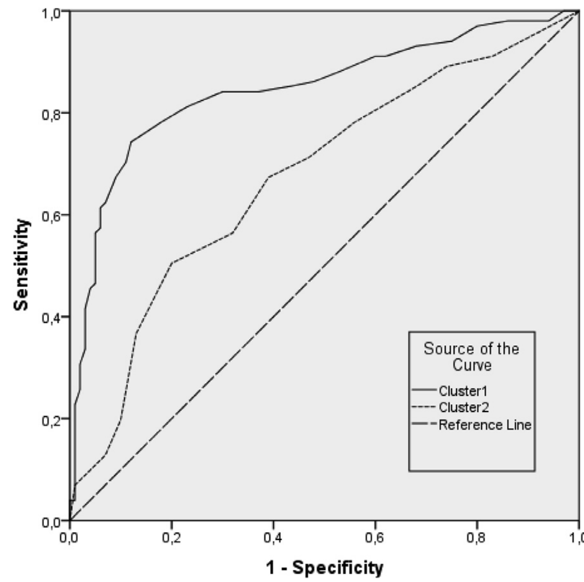

FIGURE 1. ROC curves for three subscale scores of cluster 1 and 2 of the VFI-G to discriminate voice-disordered and vocally healthy subjects.

TABLE 3.
Diagnostic Accuracy Results of the VFI-G

VFI Subscale	A_{ROC}	Threshold	Sensitivity	Specificity	LR+	LR-
Cluster 1	0.846	17.5	74.3%	88.0%	6.19	0.29
Cluster 2	0.674	7.5	50.5%	80.0%	2.52	0.62

TABLE 4.
Cluster-Related VFI-G Values of the Dysphonia and Control Groups

Characteristic		Dysphonia Group		Control Group	
		Cluster 1	Cluster 2	Cluster 1	Cluster 2
Age (correlation results)		r = 0.035	r = −0.199*	r = 0.044	r = −0.006
Sex (mean ± SD)	Male	24.39 ± 15.43	5.71 ± 3.63	9.97 ± 7.33	4.76 ± 3.61
	Female	26.40 ± 12.15	7.07 ± 3.37	10.42 ± 8.28	4.39 ± 3.37
Voice disorder (mean ± SD)	Functional dysphonia	25.84 ± 11.53	7.51 ± 3.23*,†	None	
	Laryngeal neurological dysphonia	29.25 ± 11.42	6.56 ± 2.66		
	Inflammation, epithelial lesion	24.63 ± 17.09	5.63 ± 4.11		
	Neoplasm	27.14 ± 17.73	3.29 ± 3.40*,†		
	paradoxal vocal fold movement	20.00 ± 20.66	3.67 ± 4.73		
	Larynxtrauma external	23.50 ± 17.68	6.50 ± 0.71		
Level of Vocal Usage (mean ± SD)	Elite vocal performer	16.00 ± 11.75	7.00 ± 4.16	8.93 ± 6.27	4.33 ± 3.39
	Professional voice user	29.53 ± 15.60	8.67 ± 3.02*,‡	13.94 ± 11.60	5.65 ± 3.39
	Nonvocal professional	24.47 ± 13.18	6.59 ± 3.73	8.44 ± 6.10	3.75 ± 3.03
	Nonvocal nonprofessional	22.39 ± 13.53	5.26 ± 3.56*,‡	10.78 ± 7.70	4.69 ± 3.74

* *P*-value significant at the 0.05 level.
† Significant differences between functional dysphonia and neoplasm.
‡ Significant differences between professional voice users and nonvocal nonprofessionals.

either cluster (all *P*-values > 0.05). The one-way ANOVA was therefore used. Again, sex had no impact on the values of cluster 1 ($P = 0.796$) or cluster 2 ($P = 0.632$).

VFI-G clusters related to type of voice disorder

Different types of voice disorder were investigated in the dysphonia group without reference to the control group. Because psychogenic dysphonia and dysphonia after radiotherapy were represented by only one subject, they were excluded from further analyses. Levene's test indicated no heterogeneity of VFI-G variances for either cluster so a one-way ANOVA was calculated. Voice disorder type had no impact on cluster 1 ($P = 0.868$). Cluster 2, however, showed significant differences regarding the type of voice disorder ($P = 0.017$). A post-hoc test only showed differences between functional dysphonia and neoplasms comprising of laryngeal papillomatosis, leukoplakia of larynx, and laryngeal carcinoma ($P = 0.032$).

VFI-G clusters related to the levels of vocal usage

Levene's test did not show heterogeneity of variance in either cluster for the dysphonic or control group so ANOVAs were calculated. No effect of the level of vocal usage was found on cluster 1 by the ANOVA of the dysphonic group ($P = 0.255$). Between-group differences were found ($P = 0.029$) for cluster 2 of the dysphonic group, but the post-hoc test showed significant differences only between the fatigue values of professional and nonprofessional voice users ($P = 0.018$). The levels of vocal usage had no impact on clusters 1 ($P = 0.118$) or 2 ($P = 0.312$) of the control group.

DISCUSSION

Our study confirmed, as expected, that subjects with voice disorders scored significantly higher in the two clusters (tiredness and avoidance plus physical discomfort (cluster 1), symptom improvement with rest (cluster 2)) of the VFI-G than subjects who are vocally healthy. The reconstructed VFI-G therefore seems to be more effective than the previous version in identifying vocal fatigue symptoms in people with voice disorders.

Our study is the first cross-validated study to determine thresholds for the new clusters of VFI after restructuring. A threshold of ≥ 17.5 was calculated for cluster 1, with a high specificity of 88.0% and moderate sensitivity of 74.3%. Due to the restructuring of the clusters, a direct comparison was not possible because the English version did not include threshold reports. However, although the newly validated cluster 1, developed from factors 1 and 2 of the previous version, showed a slight decrease in diagnostic accuracy compared with the previous factor 1 (sensitivity 76.2% and specificity 90.0%),[24] a significant improvement in these validity parameters was evident in comparison with the previous factor 2 (71.5% sensitivity and 81.0% specificity).[24]

The better validity scores found for cluster 1 within the reconstructed VFI-G highlight the benefit of this approach for the assessment of vocal fatigue and support its use instead of the single factors 1 and 2 of the original version. No restructuring took place when forming cluster 2 from the original factor 3, so no change in validity scores was found from the previous German cross-validation study. The previously-established threshold ≤ 7.5 for factor 3, now cluster 2, was almost identical to that found on analyses of some other languages, which ranged from ≤ 7.0[21] to ≤ 7.5.[23] Despite a good specificity of 80%, the sensitivity for cluster 2 was only approximately 50 % and thus insufficiently accurate for acceptable diagnostic use.

A lower score in cluster 2 must be interpreted as indicating increased fatigue without an improvement in vocal fatigue with rest.[21] The recovery mechanisms involved in vocal fatigue, therefore, seem difficult to measure in this manner and could explain the lower diagnostic accuracy of cluster 2. At least two scenarios are possible in which a lower score could be obtained in cluster 2, and these scenarios lead to conflicting interpretations. Firstly, individuals with or without voice-disorders but no perceived vocal fatigue do not require recovery and they, therefore, receive a lower score in cluster 2 (ie, cluster 2 has no significance for these individuals, because if no physical or psychological symptoms are perceived, physical recovery cannot be assessed). Secondly, in cases where vocal fatigue is confirmed in cluster 1 and a low level of recovery is measured in cluster 2, the information from cluster 2 is itself valuable. Future research should therefore focus on better discrimination of different recovery scenarios for vocal fatigue if cluster 2 is to count in the evaluation of VFI.

This study also examined the effects of age, gender, type of voice disorder and level of voice use on the two VFI-G clusters. These results are discussed below. The control group showed no abnormality in these variables so the discussion of these issues only covers subjects who are voice-impaired.

Age seems to have an almost negligible effect on both clusters and groups. However, there is a weak but significant association between age and cluster 2 of the dysphonia group (ie, recovery from vocal fatigue symptoms seems to be worse in individuals who are voice-impaired with increasing age). The complex conglomerate of biological processes which constitute the aging process in adults lead to physiological changes in the structure and function of the vocal apparatus with concomitant deleterious effects on the voice and its ability to regenerate.[35] For example, the following morphological discrepancies between aging and younger larynxes have been found: atrophy-related reduction in the diameter of fibers of the vocalis muscle,[36] a lower concentration of hyaluronic acid in the vocal folds,[37] reduced thickness of the vocal fold lamina propria and a decrease in epithelial cell density.[38] All of these differences may explain difficulties in recovering from vocal fatigue in older adults.

No sex differences were observed for the groups and clusters of the VFI-G. Vocal fatigue therefore seems not to differ between men and women. But doubts remain about the gender sensitivity of the VFI-G. Vocal fatigue is a common complaint among women with dysphonia, especially among professional voice users (eg, teachers), and a significantly higher prevalence of vocal fatigue symptoms has been reported in female than male teachers. The reasons for these gender differences, for example in VFI scores, appear to be multifaceted,[12] with numerous studies also finding higher incidences of vocal disorders in women than in men in other respects.[31,39−41] Regardless of occupation, 10% more women than men report a prolonged vocal disorder or vocal problem lasting longer than four weeks.[39] There could be several reasons for this phenomenon, such as gender-specific morphology and physiology of the vocal apparatus, hormonal or behavioral differences.[41] However, some other studies have found no gender differences.[42] For example, one longitudinal study of teachers lasting 12 years found that the gender difference in the prevalence of voice disorders was significant at the beginning but not at the end of the period. The increasing number of voice disorders reported in men over time has been postulated as one of the reasons for this finding.[43]

The types of voice disorder found in the present study included eight different groups of dysphonia, six of which we compared according to clusters 1 and 2 of the VFI-G. No differences were detected between these voice disorders with respect to vocal fatigue (cluster 1), but a significant difference was found between two groups (functional dysphonia vs. neoplasm) on the improvement of symptoms with rest (cluster 2) ($P = 0.032$). These two groups represent the extreme ends of the regeneration results for vocal fatigue. In functional dysphonia, a higher recovery rate is possible because no organic impairment inhibits the recovery process of physiological and biomechanical mechanisms. Our results support this hypothesis, in that functional voice disorders received higher values for cluster 2 than all of the groups of organic voice disorders. In contrast, patients with neoplastic voice disorders undergoing surgical and radiochemotherapy treatment show maximum impairment of the organic structures on which the severity of dysphonia depends. The resulting vocal fatigue appears to be more difficult to overcome in patients with neoplasms than in other vocal disorders.

No significant difference in the level of voice usage was found in cluster 1, but in cluster 2 there was a significant difference between the performance of professional and nonprofessional voice users ($P = 0.018$). Professional voice users had the highest rate of recovery from vocal fatigue. This group is subject to the highest vocal demands, alongside elite voice users (such as singers and actors), but management to prevent voice problems and vocal demands differ between the professional and non-professional groups. One key difference lies in the training and vocal techniques employed to reduce vocal fatigue and other voice problems: professional voice users, such as teachers, do not necessarily receive voice training as a standard part of their education, despite the fact that the voice performances required of them can take many hours every day.[44,45] For elite performers, the management of their vocal usage is the main part of their theoretical and practical training. This makes them highly sensitive to changes in their voice and more likely to seek medical help.[46−48] It may also explain why professional voice users need more rest, given the high demands placed upon their voices on a daily basis and the lower level of training they receive in vocal hygiene and techniques than elite users.

The fact that nonvocal nonprofessional voice users typically have a lower vocal load than other voice usage groups could explain their lower scores in cluster 2. The differences between high and low cluster 2 values are thereby reduced and the interpretability of the results is increasingly limited.

In summary, our findings regarding the diagnostic accuracy of cluster 1, and the fact that no differences in subject characteristics such as age, sex and vocal performance level were found, suggest that cluster 1 is sensitive as a tool for determining perceived vocal fatigue. These positive findings for the diagnostic accuracy of physiological and psychological aspects of vocal fatigue in cluster 1 are an improvement on the previous version of the VFI. Cluster 2 may function as an indicator of some minor changes in subject characteristics, but further optimization is necessary in order to improve its diagnostic accuracy. Furthermore, future studies should also consider cross referencing VFI-G with clinician severity ratings from auditory perceptual, acoustic, aerodynamic, and laryngeal imaging evaluations.

CONCLUSION

The quality of the VFI-G for the detection of vocal fatigue has been improved by statistically clustering factors 1 and 2 of the previous version. Cluster 1 captures the physiological and psychological symptoms of vocal fatigue in a highly specific and reasonably sensitive manner, but does not distinguish between subject characteristics such as age, voice disorder and the level of vocal use of vocally impaired and vocally healthy individuals. Cluster 2 detects some minor differences in this respect. The VFI-G questionnaire is a practicable and essentially valid instrument for assessing perceived vocal fatigue.

Acknowledgment

We would like to thank the British native speaker Ross Parfitt for final correction of our English.

REFERENCES

1. Bhattacharyya N. The prevalence of voice problems among adults in the United States. *Laryngoscope*. 2014;124:2359–2362. https://doi.org/10.1002/lary.24740.
2. Kim KH, Kim RB, Hwang DU, et al. Prevalence of and sociodemographic factors related to voice disorders in South Korea. *J Voice*. 2016;30. 246.e1-7Doi: 10.1016/j.jvoice.2015.04.010.
3. Lyberg-Åhlander V, Rydell R, Fredlund P, et al. Prevalence of voice disorders in the general population, based on the Stockholm Public Health Cohort. *J Voice*. 2019;33:900–905. https://doi.org/10.1016/j.jvoice.2018.07.007.

4. Angelillo M, Di Maio G, Costa G, et al. Prevalence of occupational voice disorders in teachers. *J Prev Med Hyg*. 2009;50:26–32.
5. Bermúdez de Alvear RM, Barón FJ, Martínez-Arquero AG. School teachers' vocal use, risk factors, and voice disorder prevalence: guidelines to detect teachers with current voice problems. *Folia Phoniatr Logop*. 2011;63:209–215. Doi: 10.1159/000316310.
6. Munier C, Kinsella R. The prevalence and impact of voice problems in primary school teachers. *Occup Med (Lond)*. 2008;58:74–76. Doi: 10.1093/occmed/kqm104.
7. Sliwinska-Kowalska M, Niebudek-Bogusz E, Fiszer M, et al. The prevalence and risk factors for occupational voice disorders in teachers. *Folia Phoniatr Logop*. 2006;58:85–101. https://doi.org/10.1159/000089610.
8. Hocevar-Boltezar I. Prevalence and risk factors for voice problems in priests. *Wien Klin Wochenschr*. 2009;121:276–281. https://doi.org/10.1007/s00508-009-1163-1.
9. Devadas U, Jose N, Gunjawate D. Prevalence and influencing risk factors of voice problems in priests in Kerala. *J Voice*. 2016;30:771.e27–771.e32. https://doi.org/10.1016/j.jvoice.2015.11.008.
10. Pestana PM, Vaz-Freitas S, Manso MC. Prevalence of voice disorders in singers: systematic review and meta-analysis. *J Voice*. 2017;31:722–727. https://doi.org/10.1016/j.jvoice.2017.02.010.
11. Niebudek-Bogusz E, Śliwińska-Kowalska M. An overview of occupational voice disorders in Poland. *Int J Occup Med Environ Health*. 2013;26:659–669. https://doi.org/10.2478/s13382-013-0146-7.
12. Hunter EJ, Banks RE. Gender differences in the reporting of vocal fatigue in teachers as quantified by the Vocal Fatigue Index. *Ann Otol Rhinol Laryngol*. 2017;126:813–818. https://doi.org/10.1177/0003489417738788.
13. Ilomäki I, Kankare E, Tyrmi J, et al. Vocal fatigue symptoms and laryngeal status in relation to vocal activity limitation and participation restriction. *J Voice*. 2017;31:248.e7–248.e10. https://doi.org/10.1016/j.jvoice.2016.07.025.
14. Welham NV, Maclagan MA. Vocal fatigue: current knowledge and future directions. *J Voice*. 2003;17:21–30. https://doi.org/10.1016/s0892-1997(03)00033-x.
15. Titze IR. *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice Hall; 1994.
16. Boucher VJ, Ayad T. Physiological attributes of vocal fatigue and their acoustic effects: a synthesis of findings for a criterion-based prevention of acquired voice disorders. *J Voice*. 2010;24:324–336. https://doi.org/10.1016/j.jvoice.2008.10.001.
17. Hunter EJ, Cantor-Cutiva LC, van Leer E, et al. Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. *J Speech Lang Hear Res*. 2020;63:509–532. https://doi.org/10.1044/2019_JSLHR-19-00057.
18. 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. https://doi.org/10.1007/s004050000299.
19. Rodríguez-Parra MJ, Adrián JA, Casado JC. Voice therapy used to test a basic protocol for multidimensional assessment of dysphonia. *J Voice*. 2009;23:304–318. https://doi.org/10.1016/j.jvoice.2007.05.001.
20. Boominathan P, Samuel J, Arunachalam R, et al. Multi parametric voice assessment: sri ramachandra university protocol. *Indian J Otolaryngol Head Neck Surg*. 2014;66(Suppl 1):246–251. doi:10.1007/s12070-011-0460-y.
21. Nanjundeswaran C, Jacobson BH, Gartner-Schmidt J, et al. Vocal Fatigue Index (VFI): development and validation. *J Voice*. 2015;29:433–440. https://doi.org/10.1016/j.jvoice.2014.09.012.
22. Naderifar E, Moradi N, Farzadi F, et al. Cross-cultural adaptation and validation of the Vocal Fatigue Index into Persian. *J Voice*. 2019;33:947.e35–947.e41. https://doi.org/10.1016/j.jvoice.2018.07.024.
23. Athira UR, Devadas U. Adaptation and validation of Vocal Fatigue Index (VFI) to Malayalam language. *J Voice*. 2020;34:810.e19–810.e24. https://doi.org/10.1016/j.jvoice.2019.03.002.
24. Barsties v. Latoszek B, Auner M, Graf S. Cross-cultural adaption and validation of the Vocal Fatigue Index in German. *J Voice*. 2021;35:161.e1–161.e13. https://doi.org/10.1016/j.jvoice.2019.08.003.
25. Contreras-Regatero S, Vila-Rovira J, Verdejo C. Validity and reliability of Spanish version of two questionnaires of vocal fatigue in female teachers. *J Voice*. 2021;35:159.e1–159.e9. https://doi.org/10.1016/j.jvoice.2019.07.008.
26. Bonetti A, Bonetti L, Čipčić O. Self-assessment of vocal fatigue in muscle tension dysphonia and vocal nodules: a preliminary analysis of the discriminatory potential of the Croatian Adaptation of the Vocal Fatigue Index (VFI-C) [published online ahead of print, 2019 Oct 9]. *J Voice*. 2019. https://doi.org/10.1016/j.jvoice.2019.08.028. S0892-1997(19)30222-X.
27. Srinivas S, Mahalingam S, Boominathan P. Cultural and linguistic adaptation of Vocal Fatigue Index for Tamil-Speaking Population (VFI-T) [published online ahead of print, 2020 Feb 21]. *J Voice*. 2020. https://doi.org/10.1016/j.jvoice.2020.01.002. S0892-1997(19)30452-7.
28. Şirin S, Öğüt MF, Bilgen C. Reliability and validity of the Turkish version of the Vocal Fatigue Index. *Turk J Med Sci*. 2020;50:902–909. https://doi.org/10.3906/sag-1908-105.
29. Zambon F, Moreti F, Ribeiro VV, et al. Vocal Fatigue Index: validation and cut-off values of the Brazilian version [published online ahead of print, 2020 Jul 18]. *J Voice*. 2020. https://doi.org/10.1016/j.jvoice.2020.06.018. S0892-1997(20)30235-6.
30. Nanjundeswaran C, van Mersbergen M, Morgan K. Restructuring the Vocal Fatigue Index using Mokken scaling: insights into the complex nature of vocal fatigue. *J Voice*. 2019;33:110–114. https://doi.org/10.1016/j.jvoice.2017.09.008.
31. De Bodt M, Van den Steen L, Mertens F, et al. Characteristics of a dysphonic population referred for voice assessment and/or voice therapy. *Folia Phoniatr Logop*. 2015;67:178–186. https://doi.org/10.1159/000369339.
32. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. 2nd Ed John Wiley & Sons, Inc.; 2000:156–164.
33. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*. 1994;271:703–707. https://doi.org/10.1001/jama.271.9.703.
34. Field A. *Discovering Statistics Using SPSS*. Sage Publications Ltd; 2009.
35. Sataloff RT, Rosen DC, Hawkshaw M, et al. The aging adult voice.. *J Voice*. 1997;11:156–160. https://doi.org/10.1016/s0892-1997(97)80072-0.
36. Martins RH, Benito Pessin AB, Nassib DJ, et al. Aging voice and the laryngeal muscle atrophy. *Laryngoscope*. 2015;125:2518–2521. https://doi.org/10.1002/lary.25398.
37. Branco A, Rodrigues SA, Fabro AT, et al. Hyaluronic acid behavior in the lamina propria of the larynx with advancing age.. *Otolaryngol Head Neck Surg*. 2014;151:652–656. https://doi.org/10.1177/0194599814544673.
38. Ximenes Filho JA, Tsuji DH, do Nascimento PH, et al. Histologic changes in human vocal folds correlated with aging: a histomorphometric study. *Ann Otol Rhinol Laryngol*. 2003;112:894–898. https://doi.org/10.1177/000348940311201012.
39. Roy N, Merrill RM, Thibeault S, et al. Prevalence of voice disorders in teachers and the general population. *J Speech Lang Hear Res*. 2004;47:281–293. https://doi.org/10.1044/1092-4388(2004/023.
40. Cohen SM, Kim J, Roy N, et al. Prevalence and causes of dysphonia in a large treatment-seeking population. *Laryngoscope*. 2012;122:343–348. https://doi.org/10.1002/lary.22426.
41. Hunter EJ, Tanner K, Smith ME. Gender differences affecting vocal health of women in vocally demanding careers. *Logoped Phoniatr Vocol*. 2011;36:128–136. https://doi.org/10.3109/14015439.2011.587447.
42. Greve K, Bryn EK, Simberg S. Voice disorders and impact of voice handicap in Norwegian student teachers. *J Voice*. 2019;33:445–452. https://doi.org/10.1016/j.jvoice.2018.01.019.

43. Simberg S, Sala E, Vehmas K, et al. Changes in the prevalence of vocal symptoms among teachers during a twelve-year period. *J Voice*. 2005;19:95–102. https://doi.org/10.1016/j.jvoice.2004.02.009.
44. Richter B, Nusseck M, Spahn C, et al. Effectiveness of a voice training program for student teachers on vocal health. *J Voice*. 2016;30:452–459. https://doi.org/10.1016/j.jvoice.2015.05.005.
45. Nusseck M, Immerz A, Spahn C, et al. Long-term effects of a voice training program for teachers on vocal and mental health [published online ahead of print, 2019 Dec 18]. *J Voice*. 2019. https://doi.org/10.1016/j.jvoice.2019.11.016. S0892-1997(19)30338-8.
46. Rosen CA, Murry T. Voice handicap index in singers. *J Voice*. 2000;14:370–377. https://doi.org/10.1016/s0892-1997(00)80082-x.
47. Cohen SM, Jacobson BH, Garrett CG, et al. Creation and validation of the singing Voice Handicap Index. *Ann Otol Rhinol Laryngol*. 2007;116:402–406. https://doi.org/10.1177/000348940711600602.
48. Murry T, Zschommler A, Prokop J. Voice handicap in singers. *J Voice*. 2009;23:376–379. https://doi.org/10.1016/j.jvoice.2007.10.013.