Cross-Cultural Adaption and Validation of the Vocal Fatigue Index in German

*,§Ben Barsties v. Latoszek, †Margrit Auner, and ‡Simone Graf, *Düsseldorf, §Dormagen, †Nümbrecht, and ‡Munich, Germany

Summary: Objective. The purpose of the study was to implement a cross-validation and adaption of the Vocal Fatigue Index (VFI) in German language.

Methods. The translation process for the German version (GV) of the VFI passed through three steps of translation before a final version was completed. Study subjects included 100 vocally healthy subjects and 101 voice-disordered subjects with various types of dysphonia. The internal consistency was determined using Cronbach's alpha (cron α) and the item-total analysis. Test-retest reliability was measured with the Pearson correlation coefficient. To assess the validity, the independent sample t test, the receiver-operating characteristic curve, the likelihood ratios, and Youden Index were used.

Results. The internal consistency across all three domains were good (#1: cron $\alpha = 0.945$, #2: cron $\alpha = 0.904$, and #3: cron $\alpha = 0.871$) and no item of the GV of the VFI had to be deleted for further analysis relating to the item-total analysis. The test-retest reliability was high to very high ($r = 0.86-0.93$). Significant higher scores were revealed in voice-disordered subjects in comparison with vocally healthy subjects in all three domains (all P values <0.01). Thresholds for the three domains of the GV of the VFI were determined at ≥ 15.5 (76.2% sensitivity and 90.0% specificity) for #1, ≥ 2.5 (71.5% sensitivity and 81.0% specificity) for #2, and ≤ 7.5 (50.5% sensitivity and 80.0% specificity) for #3.

Conclusions. It can be considered that the VFI is a valid and reliable tool identifying vocal fatigue symptoms and its severity in the German-speaking population.

Key Words: Vocal Fatigue−Self-Perception−Translating−Questionnaire−Voice−Dysphonia.

INTRODUCTION

Patient self-perception has been an integral part of the medical voice examination for many years in addition to the assessments of voice quality judgment, videostroboscopy, acoustics, and aerodynamics.[1] Self-perception with standardized questionnaires evaluate, from the patient's point of view, the presence, degree, and progression of their dysphonia. Dysphonia is characterized by changes in voice quality,[2] voice effort and vocal capacity,[3,4] vocal performance of pitch and loudness,[5−8] and sensations of the throat (eg, fatigue, burning, tight, dry, aching, tickling, sore, irritable, and a lump in the throat).[9,10] In Germany, there are some international, standardized, and validated voice questionnaires available such as the Voice Handicap Index,[11] Singing Voice Handicap Index,[12,13] Voice-related Quality of Life,[14] Vocal Tract Discomfort,[15] and Transsexual Voice Questionnaire for Male-to-Female Transsexuals.[16] However, there is currently no questionnaire for the German population to assess vocal fatigue in subjects with dysphonia. Vocal fatigue is defined as a complex multifactorial phenomenon in clinical voice practice, in which the professional voice user reports vocal tiredness and discomfort during continuous phonation while the phonatory function is reduced.[17] The process of physiological and biomechanical changes (eg, neuromuscular fatigue, increased vocal fold viscosity, reduced blood circulation, and nonmuscular tissue strain) might lead to vocal fatigue as well.[18−20] Prevalent data suggest that vocal fatigue plays an important role developing or maintaining voice disorders.[17,19] Furthermore, vocal fatigue is recognized in various occupations such as teachers, singers, actors, telemarketers, clergy, and military personnel.[19,21] The professional voice user experiences vocal fatigue provoking vocal loading through prolonged voice use, continuous and loud speaking habits or duties, adverse acoustical environments such as, for example, high background noise.[22,23] Vocal fatigue can change the laryngeal structure/function or phonatory performance of videolaryngostroboscopic findings,[17,19,24] aerodynamic measures,[17,24,25] the Dysphonia Severity Index,[17] various sensations in the larynx,[9,22,26] and acoustic measures.[9,19]

The objective quantification assessments in a present day clinical setup are inaccurate in clearly identifying/are unable to clearly identify vocal fatigue in professional voice users (eg, conventional acoustic measures[27,28] and vocal loading tasks[29,30]). Another dilemma is the lack of a uniform definition of vocal fatigue. The current state of the art of a vocal fatigue definition is mainly based on the prevailing symptoms.[9] Thus, self-perception might be still necessary to assess vocal fatigue. The development of a psychometrically disorder-specific evaluated questionnaire of self-perception yields important information of the characteristics and emphasis on vocal fatigue relating to a voice disorder. In

this context, a new questionnaire, called the Vocal Fatigue Index (VFI), has been developed. This questionnaire has shown high results in accuracy and reliability in the assessment of vocal fatigue.[31] The VFI is a self-assessment tool based on self-reported symptoms of vocal fatigue. It contains 19 items across three domains or subscales (ie, #1 = tiredness and avoidance of voice use, #2 = physical discomfort with voice use, and #3 = improvement of symptoms or lack thereof with rest) to quantify the amount of vocal fatigue one experiences. This questionnaire was originally tested by 200 voice-disordered subjects with various types of dysphonia. The VFI is scored on a five-point Likert scale (ie, 0 = never, 1 = almost never, 2 = sometimes, 3 = almost always, 4 = always). The statements of VFI are negatively worded for domains #1 and #2. The higher the score on the scale of these two domains, the greater the vocal fatigue. Domain #3 of the VFI is positively worded and, thus, a lower score indicates greater vocal fatigue. The VFI does not have a total score; however, the calculation of the VFI is based on individual scores for three domains. The threshold of the original VFI was determined for domain #1 with ≥24, for domain #2 with ≥7, and for domain #3 with ≤7.

The VFI has been developed in the English language and the VFI was evaluated in Persian[32] and Malayalam[33] languages too. In all languages, the VFI was reliable and valid to assess vocal fatigue from the view of voice-disordered subjects in comparison with vocally healthy subjects.[31−33] Accounting for the increased importance of vocal fatigue in dysphonia as reported earlier, there is a need for the German population as well, but the VFI questionnaire is not currently adapted in the German language. Thus, the objective of the present study is to translate and validate the English version of VFI into the German language, which can be used in voice protocols to evaluate the presence, the degree, and the progression of vocal fatigue in voice-disordered subjects.

METHODS

Translation process of the German version of the VFI

Prior permission was obtained from Elsevier using the statements from the original version of the VFI[31] for German translation (order number: 4581820341870). In the translation process for the German-language version of the VFI questionnaire, several specialists were involved such as an otolaryngologist (ie, specialized on voice and swallowing disorders), two speech-language pathologists (ie, both active in voice clinics), the author M.A. (ie, speech-language pathologist), an English-speaking Ph.D. student (ie, English language and literature) who is a native German, and a native English-speaker who is a medical writer with sufficient German knowledge. Furthermore, the standards of linguistic environment for the translation process were followed. The main focus was on the content validity and linguistic quality of the translation. Particularly, the correct use of the technical terminology and the equivalence of the items compared to the original version were considered. A back-translation by the native

English speaker of the translation was achieved to avoid inter-language interferences.[34]

Table 1 shows the process of the German translation of the VFI. The questionnaire was independently translated into German by all experts. Before a consensus German version (GV) of the items was reached, a bilingual German-English speech-language pathologist checked the intelligibility and terminology of the German-language version of the VFI to ensure the minimization of semantic differences.[35,36] The final GV of the VFI was back-translated and compared with the original version of the VFI. The back-translation of the original English version of the VFI showed high-coverage equivalence of the items in a quantitative single-word analysis (ie, 95.8% of the total word material) and partly verbal consistency of the formulations (ie, 52, 6% of the total word material) in comparison to the original VFI version. The content validity and semantic consistency of the final translated GV of the VFI met the requirements for further analyses. The final translated German VFI has a new layout and some demographic data (eg, age, gender, occupation, and voice diagnosis) were adjusted in comparison to the original VFI version (see Supplementary Material).

Subjects

The GV of the VFI was selected in two ways: (1) as an online version in a computer-aided web interview mode for the group of healthy-voiced volunteers (control group), and (2) the personal interview mode for the group of subjects with dysphonia (dysphonia group). The participation was anonymous and was carried out on voluntary basis throughout Germany; the subjects were given no explanation for completing the questionnaire. The inclusion criteria for both groups were that all subjects were at least 18 years old. The subjects of the control group had to consider their voice as normal. They had to be free of voice problems, no history of chronic laryngeal diseases or other long-lasting voice disorders. Furthermore, they had to confirm that they had no voice complaints, no voice therapy, and no acute laryngitis for at least 3 months. Finally, all of them were free from any known hearing impairments and free from common cold or upper respiratory infections at the time of the survey. The dysphonia group could present various organic and nonorganic etiologies and various degrees in dysphonia severity diagnosed by an otolaryngologist. Subjects with a diagnosis of an acute laryngitis, chronic respiratory disorders, chronic obstructive pulmonary disease, chronic respiratory deficits, or neurological diseases (eg, amyotrophic lateral sclerosis, Parkinson's disease, and multiple sclerosis) were excluded.

The data were collected at two measuring times because of test-retest reliability of the VFI for both groups. In the dysphonia group, 183 questionnaires were collected over a period of 5 months and 81 had to be excluded due to missing retest information. In the control group, a total of 272 online questionnaires were answered and 172 had to be excluded due to missing retest or personal information.

TABLE 1.
Process of Translation and Cultural Adaptation to a German Version of the Vocal Fatigue Index (VFI)

Statement	Original English Version	Translation Into German	Back-Translation of the GV Into English (T6)
1	I do not feel like talking after a period of voice use.	T1: Ich habe keine Lust mehr zu reden, wenn ich längere Zeit gesprochen habe. T2: Nach einer Phase der Sprechanstrengung bevorzuge ich es, nicht mehr zu sprechen. T3: Nach einer Zeit, in der ich meine Stimme beansprucht habe, ist mir nicht danach zu reden. T4: Mir ist nicht nach Reden zumute nachdem ich eine Weile gesprochen habe. T5: Nach längerem Stimmgebrauch habe ich keine Lust mehr zu reden. GV: Wenn ich länger gesprochen habe, ist mir nicht mehr nach reden zumute.	When I have spoken for a longer time, I do not feel like talking anymore.
2	My voice feels tired when I talk more.	T1: Meine Stimme ist erschöpft, wenn ich mehr spreche. T2: Wen ich viel spreche bemerke ich, dass meine Stimme müde wird/bemerke ich die Stimmermüdung. T3: Meine Stimme ermüdet, wenn ich viel spreche. T4: Meine Stimme wird müde, je mehr ich rede. T5: Meine Stimme wird müde, wenn ich mehr spreche. GV: Meine Stimme ermüdet, wenn ich viel spreche.	My voice becomes tired when I speak a lot.
3	I experience increased sense of effort with talking.	T1: Ich habe das Gefühl von Anstrengung wenn ich länger spreche. T2: Ich erlebe, dass ich mich beim Sprechen anstrengen muss. T3: Es beansprucht mich zu sprechen. T4: Ich verspüre erhöhte Anstrengung wenn ich rede. T5: Ich habe das Gefühl mich beim Sprechen mehr anzustrengen. GV: Ich habe das Gefühl von erhöhter Anstrengung beim Sprechen.	While speaking I have a feeling of higher exertion.
4	My voice gets hoarse with voice use.	T1: Meine Stimme wird heiser beim Sprechen. T2: Meine Stimme wird rau/heiser wenn ich spreche. T3: Wenn ich meine Stimme nutze, werde ich heiser.	My voice becomes hoarse when I speak.

(*Continued*)

TABLE 1. (*Continued*)

Statement	Original English Version	Translation Into German	Back-Translation of the GV Into English (T6)
		T4: Meine Stimme wird heiser, wenn ich sie benutze.	
		T5: Wenn ich meine Stimme nutze, wird sie heiser.	
		GV: Meine Stimme wird heiser wenn ich spreche.	
5	It feels like work to use my voice.	T1: Es fühlt sich wie Arbeit an meine Stimme zu benutzen.	It feels like hard work when I use my voice.
		T2: Es fühlt sich wie Arbeit an, wenn ich meine Stimme nutze.	
		T3: Meine Stimme zu nutzen fühlt sich wie Arbeit an.	
		T4: Es fühlt sich wie Arbeit an, wenn ich meine Stimme benutze.	
		T5: Es kommt mir wie Arbeit vor, meine Stimme zu benutzen.	
		GV: Es fühlt sich wie Arbeit an, wenn ich meine Stimme nutze.	
6	I tend to generally limit my talking after a period of voice use.	T1: Wenn ich länger gesprochen habe, versuche ich danach generell weniger zu sprechen.	I tend to speak less after I have already spoken for a while.
		T2: Ich tendiere/neige dazu, meine Stimme zu schonen, wenn ich eine Weile mehr gesprochen habe.	
		T3: Nach einer Zeit, in der ich meine Stimme benutzt habe, versuche ich meist weniger zu sprechen.	
		T4: Nachdem ich eine Weile meine Stimme benutzt habe, neige ich dazu, meine Stimmnutzung einzugrenzen.	
		T5: Generell neige ich dazu, meinen Gesprächsumfang einzuschränken, nachdem ich eine Weile gesprochen habe.	
		GV: Ich neige dazu, meinen Gesprächsumfang einzuschränken, nachdem ich eine Weile mehr gesprochen habe.	
7	I avoid social situations when I know I have to talk more.	T1: Ich vermeide soziale Situationen, bei denen ich mehr sprechen muss.	I avoid meeting people when I know that I have to talk all the time.
		T2: Ich vermeide Situationen im privaten Umfeld, wenn ich wei, dass ich dort viel sprechen muss.	
		T3: Wenn ich wei, dass ich viel reden muss, vermeide ich soziale Begegnungen.	

(*Continued*)

TABLE 1. (*Continued*)

Statement	Original English Version	Translation Into German	Back-Translation of the GV Into English (T6)
		T4: Ich meide soziale Situationen, wenn ich wei, dass ich mehr reden muss.	
		T5: Ich vermeide gesellige Situationen, wenn ich wei, dass sie mit erhöhtem Sprechaufwand verbunden sind.	
		GV: Ich vermeide soziale Begegnungen wenn ich wei, dass ich dabei viel sprechen muss.	
8	I feel I cannot talk to my family after a work day.	T1: Nach einem Arbeitstag habe ich das Gefühl nicht mit meiner Familie sprechen zu können.	After a working day I feel that I cannot talk with my family anymore.
		T2: Ich habe das Gefühl, dass ich nach einem Arbeitstag nicht mehr mit meiner Familie sprechen kann.	
		T3: Nach einem (langen) Arbeitstag glaube ich nicht, dass ich noch mit meiner Familie reden kann.	
		T4: Ich fühle mich, dass ich mit meiner Familie nicht reden kann, nachdem ich einen Arbeitstag hatte.	
		T5: Am Ende eines Arbeitstages habe ich das Gefühl, dass ich mich nicht einmal mehr mit meiner Familie unterhalten kann.	
		GV: Ich habe das Gefühl, dass ich nach einem Arbeitstag nicht mehr mit meiner Familie sprechen kann.	
9	It is effortful to produce my voice after a period of voice use.	T1: Es ist anstrengend zu sprechen nach einer Zeit des Stimmgebrauchs.	It is exhausting to create voice after having spoken for a while.
		T2: Es ist anstrengend nach einer Phase der stimmlichen Anstrengung meine Stimme zu nutzen.	
		T3: Nachdem ich meine Stimme eine Weile lang benutzt habe, kostet es Kraft, noch Stimme zu erzeugen.	
		T4: Es braucht viel Anstrengung, meine Stimme weiter zu benutzen, nachdem ich sie schon eine Weile benutzt habe.	
		T5: Nach längerem Stimmgebrauch strengt es mich an, Stimme zu erzeugen.	
		GV: Es ist anstrengend Stimme zu erzeugen, wenn ich bereits eine Weile gesprochen habe.	

(*Continued*)

TABLE 1. (*Continued*)

Statement	Original English Version	Translation Into German	Back-Translation of the GV Into English (T6)
10	I find it difficult to project my voice with voice use.	T1: Ich finde es schwierig, beim Sprechen meine Stimme durchzusetzen. T2: Ich finde es schwierig, mit meiner Stimme etwas rüberzubringen. T3: Mit zunehmender Stimmnutzung finde ich es schwierig, meine Stimme zu projizieren. T4: Ich finde es schwierig, meine Stimme zu erheben, nachdem ich meine Stimme eine Weile benutzt habe. T5: Ich finde es schwierig, meine Stimme durchzusetzen beim Sprechen. GV: Ich finde es schwierig, mich beim Sprechen stimmlich durchzusetzen.	It is difficult for me to get through vocally while talking/speaking.
11	My voice feels weak after a period of voice use.	T1: Nach längerem Sprechen ist meine Stimme schwach. T2: Meine Stimme fühlt sich schwach an, wenn ich eine Weile gesprochen habe. T3: Nach einer Zeit der Stimmnutzung fühlt sich meine Stimme schwach an. T4: Meine Stimme fühlt sich schwach an, nachdem ich meine Stimme eine Weile benutzt habe. T5: Meine Stimme fühlt sich kraftlos an wenn ich sie längere Zeit benutze. GV: Meine Stimme fühlt sich schwach an wenn ich eine Weile gesprochen habe.	My voice feels weak after having spoken for a while.
12	I experience pain in the neck at the end of the day with voice use.	T1: Nach einem Tag Sprechen habe ich Nackenschmerzen. T2: Ich erlebe es, dass ich Nackenschmerzen am Ende eines sprechanstrengenden Tages habe. T3: Wenn ich einen Tag lang meine Stimme viel benutze, habe ich (abends) Schmerzen im Nacken. T4: Mein Genick tut weh am Ende des Tages an dem ich meine Stimme benutzt habe. T5: Am Ende des Tages bekomme ich Nackenschmerzen, wenn ich tagsüber gesprochen habe. GV: Ich bekomme abends Nackenschmerzen nach einem Tag mit Stimmgebrauch.	After having used my voice during the day I get neck pain in the evening.

(*Continued*)

TABLE 1. (*Continued*)

Statement	Original English Version	Translation Into German	Back-Translation of the GV Into English (T6)
13	I experience throat pain at the end of the day with voice use.	T1: Nach einem Tag Sprechen habe ich Halsschmerzen. T2: Ich erlebe es, dass ich Halsschmerzen am Ende eines sprechanstrengenden Tages habe. T3: Wenn ich einen Tag lang meine Stimme viel benutze, habe ich (abends) Halsschmerzen. T4: Mein Hals tut weh am Ende des Tages an dem ich meine Stimme benutzt habe. T5: Ich verspüre abends Halsschmerzen, wenn ich tagsüber gesprochen habe. GV: Ich habe abends Halsschmerzen, wenn ich tagsüber gesprochen habe.	After having spoken during the day I have a sore throat in the evening.
14	My voice feels sore when I talk more.	T1: Meine Stimme fühlt sich wund an, wenn ich länger spreche. T2: Meine Stimme fühlt sich schlimm an wenn ich mehr spreche. T3: Meine Stimme wird rau, wenn ich viel rede. T4: Meine Stimme fühlt sich empfindlich an, wenn ich mehr rede. T5: Mein Hals fühlt sich wund an, wenn ich mehr spreche. GV: Mein Hals fühlt sich wund an wenn ich mehr spreche.	My throat feels sore when I speak more than normal.
15	My throat aches with voice use.	T1: Mein Hals schmerzt, wenn ich spreche. T2: Mein Hals tut weh, wenn ich spreche. T3: Mein Hals tut weh, wenn ich viel rede. T4: Mein Hals fühlt sich schmerzhaft an, wenn ich meine Stimme benutze. T5: Mein Hals tut weh, wenn ich spreche. GV: Mein Hals tut weh wenn ich spreche.	My throat hurts when I speak.
16	I experience discomfort in my neck with voice use.	T1: Ich habe Nackenschmerzen beim Sprechen. T2: Mein Nacken fühlt sich nicht gut an, wenn ich spreche. T3: Mein Hals fühlt sich unbehaglich an, wenn ich viel rede. T4: Ich erlebe Beschwerden in meinem Genick, wenn ich meine Stimme benutze. T5: Ich empfinde Beschwerden im Nackenbereich bei Stimmgebrauch. GV: Ich empfinde Beschwerden im Nackenbereich wenn ich spreche.	I feel pain in my neck when I speak.

(*Continued*)

TABLE 1. (*Continued*)

Statement	Original English Version	Translation Into German	Back-Translation of the GV Into English (T6)
17	My voice feels better after I have rested.	T1: Meine Stimme ist besser nach einer Erholungsphase. T2: Meine Stimme ist besser, nachdem ich mich ausgeruht habe T3: Meine Stimme wird besser, nachdem ich Pause gemacht habe. T4: Meine Stimme fühlt sich besser an, nachdem ich mich ausgeruht habe. T5: Meiner Stimme geht es besser, nachdem ich mich ausgeruht habe. GV: Meine Stimme fühlt sich besser an nachdem ich mich ausgeruht habe.	My voice feels better after I have relaxed.
18	The effort to produce my voice decreases with rest.	T1: Die Stimmanstrengung sinkt nach einer Erholungsphase. T2: Die Anstrengung, die ich zur Stimmproduktion brauche, nimmt ab, wenn ich mich mehr ausruhe. T3: Durch Pausen wird die Anstrengung, meine Stimme zu benutzen, weniger. T4: Ich brauche weniger Mühe zu sprechen, wenn ich mich ausruhe. T5: Die Anstrengung im Stimmgebrauch wird geringer durch Ruhephasen. GV: Die Anstrengung bei der Stimmproduktion wird geringer durch Ruhephasen.	The exertion in voice production becomes less by taking breaks.
19	The hoarseness of my voice gets better with rest.	T1: Die Heiserkeit wird nach einer Erholungsphase besser. T2: Die Rauigkeit/Heiserkeit meiner Stimme wird besser, wenn ich mich mehr ausruhe. T3: Meine Stimme ist weniger rau, wenn ich Pausen mache. T4: Die Heiserkeit meiner Stimme verbessert sich, wenn ich mich ausruhe. T5: Die Heiserkeit meiner Stimme wird geringer, wenn ich mich ausruhe. GV: Die Heiserkeit meiner Stimme lässt nach durch Ruhephasen.	The hoarseness of my voice becomes less by taking breaks.

T1, English-German translator number one; T2, English-German translator number two; T3, English-German translator number three; T4, English-German translator number four; T5, English-German translator number five; T6, English-German translator number six; GV, German version of the compilation of translations T1–T5.

Thus, a total of 102 evaluable questionnaires for the dysphonia group and 100 evaluable questionnaires for the control group were compiled, which met the inclusion criteria. This corresponds to a return rate of 55.74% for the dysphonia group and 51.41% for the control group. The dysphonic group had a mean age of 40.8 years ± 16.8 years and 31 males (30.7%) and 70 females (69.3%) were included. In the control group, a mean age of 51.0 years ± 15.6 years was noted, in which 29 males (20%) and 71 females (80%) participated.

The dysphonia group consisted of 101 voice patients who represented a rather common and clinically discriminative group of voice disorders, as listed in Table 2. The preliminary diagnosis was based on clinical examination (ie, complaints, history) and results of laryngoscopy. The voice-disordered subjects were mostly recruited from several speech-language pathology practices who had an otolaryngology letter of referral, in which the equipment of the otolaryngologists for the diagnosis were unknown. Furthermore, Table 2 reports further demographic data of the dysphonia group next to the diagnosis. The incidence of voice disorders in our series is considered to be clinically representative for the population of voice-disordered subjects reflecting different age and gender groups, as well as different types and severity levels of voice quality as reported in other studies.[37−39]

Statistics

Statistical analyzes were completed using *SPSS*, version 23 for Windows (IBM Corp., Armonk, NY). First, the Cronbach's alpha (cron α) coefficient was calculated for the analysis of the reliability by means of the internal consistency of the total scale (19 items). Interpretation guidelines for cron α shows acceptable value ranging from 0.70 to 0.95.[40,41] A lower cron α indicates poor interrelatedness between items or heterogeneous construct. A higher cron α suggests that items are redundant. Additionally, the item-total analysis

was conducted to investigate the corrected item-total correlation and the cron α if the item was deleted. The following interpretation guidelines were used deleting an item of the questionnaire if two aspects were present: (1) the item should correlate with the total of $r \leq 0.20$,[41] and (2) the cron α is higher if the item is deleted.

Second, the test-retest reliability was determined with the Pearson correlation coefficient. Interpretation guidelines for r were provided by Frey et al[42]: $r < 0.20$ slight correlation, $r = 0.20−0.40$ low correlation, $r = 0.41−0.70$ moderate correlation, $r = 0.71−0.90$ high correlation, $r > 0.90$ very high correlation.

Third, a t test for independent samples was prepared for both groups to verify whether there was acceptable diagnostic discrimination in the three domains of the VFI.

Fourth, a receiver-operating characteristic (ROC) curve with regard to sensitivity (ie, the recording of the actually voice-disordered subjects) and specificity (ie, the recording of the actually vocally healthy subjects) was conducted to analyze diagnostic accuracy. To estimate the power of VFI in discriminating between vocally healthy and voice-disordered subjects, the area under the ROC curve (A_{ROC}) was used. An A_{ROC} of >0.9 is considered to be exceptionally good, while a measure of <0.7 is considered to be low but an $A_{ROC} = 0.5$ corresponds with chance-level diagnostic accuracy.[43,44] In order to find the optimal threshold, the Youden Index was calculated providing the maximum of sensitivity + specificity − 1. To verify the applicability of the best threshold for a clinical decision, the likelihood ratio (LR) was conducted. The LR assesses the balance between the "LR for a positive result" (LR+) and the "LR for a negative result" (LR−), which are defined as the sensitivity/ (1 − specificity), and (1 − sensitivity)/specificity, respectively. A general guideline for LR statistics was provided. The results of LR+ >10 and LR− <0.1 generate large and often conclusive diagnostic evidence, LR+ = 5−10 and LR− = 0.1−0.2 indicate moderate diagnostic evidence, LR + = 2−5 and LR− = 0.2−0.5 show small diagnostic

TABLE 2.
Demographic Data and Types/Deviation of Voice Disorders of the Dysphonia Group

| Groups | Types of Dysphonia | Number | Gender | | Age in Years Mean ± SD |
			Female	Male	
Dysphonia group	Functional dysphonia	55	46	9	46.02 ± 15.66
	Laryngeal neurological dysphonia (ie, n. vagus paralysis, n. laryngeus recurrens paralysis, and spasmodic dysphonia)	16	7	9	58.63 ± 11.21
	Inflammation, epithelial lesion	16	11	5	54.69 ± 16.30
	Neoplasm	7	3	4	62.71 ± 8.30
	Idiopathic dysphonia (paradoxal vocal fold movement)	3	1	2	59.33 ± 12.90
	Larynxtrauma external	2	0	2	36.50 ± 6.36
	Dysphonia after radiotherapy	1	1	0	72.00
	Psychogenic dysphonia	1	0	1	51.00

evidence, and LR+ = 1−2 and LR− = 0.5−1 assess as rarely relevant diagnostic evidence.[45]

All results were considered statistically significant at $P \leq 0.05$.

RESULTS

Internal consistency

Table 3 presents the results of the internal consistency measures. First, the cron α was used to assess the internal consistency of the VFI. All three domains showed good results to denote the GV of VFI with high internal consistency (#1: cron $\alpha = 0.945$, #2: cron $\alpha = 0.904$, and #3: cron $\alpha = 0.871$), because cron α values were above 0.70 and lower than 0.95.

Second, the item-total analysis revealed good results for all three domains as well. Domain #1 ranged between $r = 0.669−0.848$ and all cron α were lower than 0.945 if items were deleted from domain #1. The second domain of the VFI demonstrated a range of $r = 0.724−0.782$, in which no cron α were higher than 0.904 if items were deleted from domain #2. Also, the last domain of #3 of the VFI reached r values between $r = 0.677$ to $r = 0.829$. Only one of the three items (ie, item 19) had a slightly higher cron $\alpha = 0.886$ than the total cron α of 0.871 for domain #3. However, the correlation for item 19 was meaningfully higher with $r = 0.677$ than $r = 0.20$.[41] To exclude this item from the VFI, two criteria have to be attained from the item-total analysis but for item 19 only one of the two criteria was met for exclusion. Therefore, all items of the three domains reached the criteria as mentioned earlier and no item of the GV of the VFI had to be deleted for further analysis.

Test-retest reliability

All 201 subjects completed the VFI a second time after a week to assess the test-retest reliability. High to very dependable level of test-retest reliability was measured for the three domains: #1 = tiredness and avoidance of voice use ($r = 0.93$, $P < 0.01$), #2 = physical discomfort with voice use ($r = 0.88$, $P < 0.01$), and #3 = improvement of symptoms or lack thereof with rest ($r = 0.86$, $P < 0.01$).

Validity

The independent t test showed significant differences between the two groups of the mean scores in all three domains of GV of the VFI (all P values <0.01). The results indicated that the dysphonic group had higher scores in all domains than the controls (Table 4).

To determine the optimal threshold, Table 5 summarizes the results of the diagnostic accuracy for the GV of the VFI. In domains #1 and #2 of the VFI, the A_{ROC} demonstrated acceptable precision to differentiate between voice-disordered and vocally healthy subjects ($A_{ROC} = 0.851$ and $A_{ROC} = 0.769$). The A_{ROC} of 0.674 in domain #3 showed lower discriminating power, but the result was still high enough that no chance level of diagnostic accuracy was present.

Under consideration of the Youden Index, the optimal thresholds for the three domains plus diagnostic strength in terms of sensitivity, specificity, and LR statistics resulting from the ROC curve (Figure 1) is reported in Table 5. In domain #1 of the VFI, the threshold of ≥ 15.5 showed an acceptable sensitivity with 76.2% and a high specificity with 90.0%, in which the LR statistics revealed, moreover, moderate diagnostic evidence. The domains of #2 and #3 of the VFI scored lower in diagnostic accuracy based on the sensitivity of 71.5% and 50.5%, and specificity of 81.0% and 80.0%, respectively, with small diagnostic evidence resulting from LR statistics. The threshold for domain #2 was calculated at ≥ 2.5 and for domain #3 at ≤ 7.5.

TABLE 3.
**Internal Consistency Results for the Three Subscale Sores of the VFI**

Internal Consistency Measures	#1 Tiredness and Avoidance of Voice Use	#2 Physical Discomfort With Voice Use	#3 Improvement of Symptoms or Lack Thereof With Rest
Cronbach's alpha	0.945	0.904	0.871
Item-total analysis	$r = 0.669$ to $r = 0.848$	$r = 0.724$ to $r = 0.782$	$r = 0.677$ to $r = 0.829$

TABLE 4.
Results for the Three Subscale Sores of the VFI in the Dysphonic and Control Group

Groups	#1 Tiredness and Avoidance of Voice Use		#2 Physical Discomfort With Voice Use		#3 Improvement of Symptoms or Lack Thereof With Rest	
	Mean	SD	Mean	SD	Mean	SD
Control group	8.74	6.07	1.55	2.50	4.50	3.42
Dysphonia group	20.68	9.71	5.10	4.64	6.65	3.49

TABLE 5.
Diagnostic Accuracy Results of the VFI

VFI Subscale	A_{ROC}	Threshold	Sensitivity	Specificity	LR+	LR−
Domain #1	0.851	15.5	76.2%	90.0%	7.62	0.26
Domain #2	0.769	2.5	71.5%	81.0%	3.76	0.35
Domain #3	0.674	7.5	50.5%	80.0%	2.52	0.62
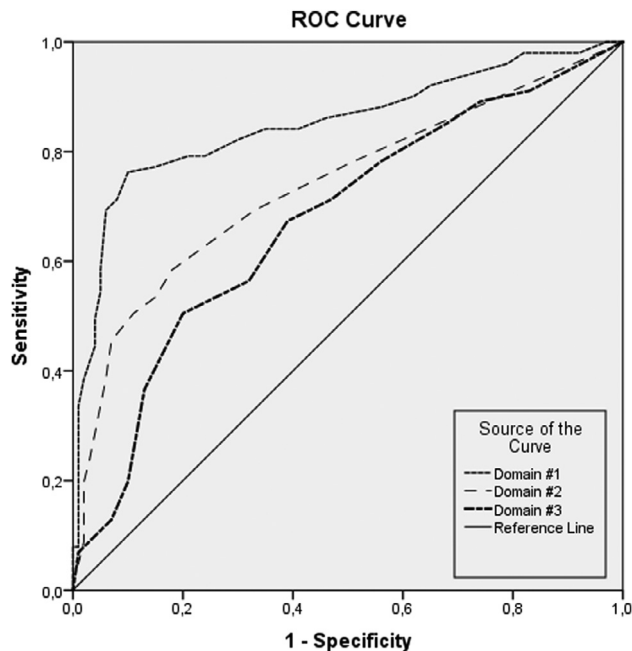
FIGURE 1. ROC curve analysis for three subscale scores of domains #1, #2, and #3 of the GV of the VFI to discriminate voice-disordered and vocally healthy subjects.

DISCUSSION

Vocal fatigue is often a symptom in voice disorders and the identification of vocal fatigue is essential at an early stage to prevent the development of dysphonia. The usefulness of assessments in the clinical evaluation of vocal fatigue is confirmed with self-perception.[9,17,22,46] The VFI is a questionnaire to assess vocal fatigue, which was tested in other languages and showed good results in reliability and validty.[31−33] The selected two groups for the English (ie, 105 subjects in the dysphonia group and 70 individuals of the control group) and Persian (ie, 80 subjects in the dysphonia group and 50 individuals of the control group) versions of the VFI were comparable to the present study.[31,32] The Malayalam version of the VFI included only primary school teachers with (ie, $n = 287$) and without self-reporting (ie, $n = 241$) of voice problems.[33] To consider teachers as an investigation group for vocal fatigue seems useful because many studies have reported that teachers have higher prominent vocal fatigue indicators than others.[47−50] However, these indicators were still lower than that of the voice-disordered population.[31] According to these differences, the VFI has potential to indicate increased vocal fatigue of voice use indicators identifying at-risk individuals at an earlier stage.[47]

The aim of the present study was to determine the cultural adaption and validation of the VFI in the German language. During the translation process, particular attention was paid to involving an appropriate number of experts (ie, six in total) in the translation process in order to avoid, as far as possible, interlanguage interference and achieve the best possible equivalence of content. The results of the translation and cultural adaption process revealed that no items of the GV of the VFI were deleted. The language deviations were low. The final consensus version of the VFI in the German language was subjected to a quantitative single-word analysis following the back-translation into English and showed a high level of coincidence regarding the substantive equivalence of the items compared to the original version (95.8% of the total word material) as well as a partial verbal agreement of the formulations (52.6% of total word material). Additionally, the internal consistency showed good results with comparable findings as reported in the other VFI validation studies of English, Persian, and Malayalam languages.[31−33]

The distribution of subjects in terms of group size turned out to be very homogeneous with the exception of gender distribution. Males were under-represented in both groups (ie, the dysphonic group with 30.7% and the control group with 20%). However, these lower numbers in males are comparable with the validation study in English[31] (33.3% in the dysphonia group and 30% in vocally healthy subjects), and Mayalayam[33] (34% in teachers with no voice problems and 20.9% in teachers with voice problems) languages.

The test-retest reliability of the GV of the VFI showed high reliability for all three domains of the VFI during time intervals. These results are closely to the reliability results of the other VFI studies.[31−33] However, only domain #3 of the English version showed a weaker test-retest reliability of $r = 0.30$ ($P < 0.01$).[31]

According to the significant differences between the dysphonic group and the control group in all three domains of the GV of the VFI, the findings suggest that the VFI can effectively identify the vocal fatigue symptoms in subjects with voice disorders.

Thresholds for the three domains of the English[31] and Malayalam[33] VFI versions were analyzed such as in the present study. The Persian VFI version did not analyze ROC statistics or comparable statistics to determine thresholds for the three domains of the VFI.[32]

For domain #1, the thresholds were ≥ 15.5, ≥ 16.5,[33] and ≥ 24,[31] in which the present study had the lowest threshold level. Specificity results were different with a higher level in

the present study (90.0% vs 71.4%), but a nearly comparable sensitivity level of 76.2% in the present study and 80.0% in the validation study of Malayalam language.[33]

For domain #2, a threshold of ≥2.5 was determined in the present study. However, this threshold is significantly lower than in other validation studies of the English (≥7.0)[31] and Malayalam (≥6.5) versions of the VFI. However, the levels of sensitivity and specificity of the present study showed acceptable results for domain #2 (71.5% and 81.0%, respectively), in which both levels were higher than in the validation study of the Malayalam language (71.1% and 70.1%).[33] According to these differences in thresholds of domain #2, the present results underline the concerns of the developers of the VFI to serve domain #2 as a good predictor of vocal fatigue.[31] There was a further investigation to determine the role of domain #2 in chronic vocal fatigue and to determine if this subscale is conceptually distinct from domain #1 using a restructuring of the VFI to develop a hierarchical understanding of the latent trait of vocal fatigue.[51] The results showed that domains #1 and #2 were combined and reordered their structure after restructuring.[51] Future studies should investigate again its validation levels and thresholds for clinical practice.

The present threshold ≤7.5 of domain #3 was nearly identical to the other languages ranging from ≤7.0[31] to ≤7.5.[33] However, the level of sensitivity and specificity differed in which the present study revealed a higher level of specificity (80.0% vs 69.9%), but a lower level of sensitivity (50.5% vs 71.1%).[33] Although the mean results significantly differed between both groups in domain #3 of the VFI, in the present study this subscale showed the lowest diagnostic accuracy results in discrimination power, sensitivity, and specificity. A lower score in this subscale indicates worse fatigue without improving vocal fatigue with rest,[31] but to measure the recovery in vocal fatigue in the present way in this questionnaire could be an issue. Because two scenarios are conceivable in obtaining a lower score in domain #3, the statement of recovery in vocal fatigue is self-contradictory. Vocally healthy or voice-disordered professional voice users without experience of vocal fatigue do not need any recovery and scored lower. In the present study, 30% of voice-disordered and 94% of vocally healthy subjects scored lower than the threshold of 15.5 of domain #1 and 2.5 of domain #2, which might have no vocal fatigue symptoms relating to tiredness, avoidance, and physical discomfort of voice use. Voice-disordered professional voice users who experienced vocal fatigue mostly score higher in domain #3, which was measured in the present study and others,[31−33] but the impact of vocal fatigue is much worse if a lower value is scored.[31] Additionally, the Voice Handicap Index, a standardized questionnaire to verify the perceived impact of voice problems,[52] had a low correlation with domain #3, although the two other domains of the VFI revealed a marked relationship with the Voice Handicap Index.[53] In summary, future investigations should consider a better discrimination in the recovery of different scenarios of vocal fatigue for professional voice users with and without vocal fatigue. Finally, in order to adapt the arch to clinical experience, further modifications of the VFI are conceivable.

There could be a modification of the questions regarding the chronic vocal fatigue symptoms, as for example in item 3 "I have CONSTANTLY the feeling of increased effort when speaking" or item 4 "My voice is ALWAYS hoarse when I speak." In addition, with such modification, an increased value might be a clearer indication of chronic vocal fatigue (ie, that the voice has had to withstand unfavorable professional, private, or environmental conditions for a longer period of time). The longer the chronic symptoms of vocal fatigue already exist, the more automated are vocal compensatory mechanisms on the glottal and physical levels are.

Future studies by using VFI to investigate the prevalent factors in vocal fatigue might be valuable as well. The focus could be to analyze, as for example, gender, age, profession, and voice disorders. A beginning was started by Hunter and Banks,[47] who investigated gender differences in teachers with vocal fatigue using the VFI. Another study established that the classroom capacity significantly affected teachers' reported amounts of vocal fatigue on the VFI.[54] Furthermore, a peak effect of increased vocal fatigue was measured occurring between 40 and 45 years in teachers in all three domains of the VFI.[54]

CONCLUSION

The GV of the VFI was determined as reliable and differentiated validly between the dysphonia group and the control group. The VFI can be easily integrated into the clinical routine, can be processed in a short time, and takes little time for the implementation and evaluation. The relatively small number of 19 items has so far proved sufficient to evaluate self-perception of vocal fatigue.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.jvoice.2019.08.003.

REFERENCES

1. Dejonckere PH, Bradley P, Clemente P, et al. Abasic 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.
2. Barsties B, De Bodt M. Assessment of voice quality: current state-of-the-art. *Auris Nasus Larynx*. 2015;42:183–188.
3. Ford Baldner E, Doll E, van Mersbergen MR. A review of measures of vocal effort with a preliminary study on the establishment of a vocal effort measure. *J Voice*. 2015;29:530–541.
4. Welham NV, Maclagan MA. Vocal fatigue: current knowledge and future directions. *J Voice*. 2003;17:21–30.
5. Ikeda Y, Masuda T, Manako H, et al. Quantitative evaluation of the voice range profile in patients with voice disorder. *Eur Arch Otorhinolaryngol*. 1999;256(suppl 1):S51–S55.
6. Ma E, Robertson J, Radford C, et al. Reliability of speaking and maximum voice range measures in screening for dysphonia. *J Voice*. 2007;21:397–406.

7. D'Alatri L, Marchese MR. The speech range profile (SRP): an easy and useful tool to assess vocal limits. *Acta Otorhinolaryngol Ital*. 2014;34:253–258.

8. Caffier PP, Möller A, Forbes E, et al. The vocal extent measure: development of a novel parameter in voice diagnostics and initial clinical experience. *Biomed Res Int*. 2018;2018: 3836714.

9. Solomon NP. Vocal fatigue and its relation to vocal hyperfunction. *Int J Speech Lang Pathol*. 2008;10:254–266.

10. Mathieson L, Hirani SP, Epstein R, et al. Laryngeal manual therapy: a preliminary study to examine its treatment effects in the management of muscle tension dysphonia. *J Voice*. 2009;23:353–366.

11. Nawka T, Wiesmann U, Gonnermann U. Validation of the German version of the Voice Handicap Index. *HNO*. 2003;51:921–930.

12. Lorenz A, Kleber B, Büttner M, et al. Validation of the German version of the Singing Voice Handicap Index. *HNO*. 2013;61:699–706.

13. Barsties B, Kropp J, Dicks P, et al. Reliability and validity of the "Voice Handicap Index (VHI) adapted to the singing voice". *Laryngorhinootologie*. 2015;94:441–446.

14. Schwanfelder C, Eysholdt U, Rosanowski F, et al. Voice-related quality-of-life: structure, validity and factors of the German questionnaire. *Folia Phoniatr Logop*. 2008;60:241–248.

15. Lukaschyk J, Brockmann-Bauser M, Beushausen U. Transcultural adaptation and validation of the German version of the vocal tract discomfort scale. *J Voice*. 2017;31. 261.e1−261.e8.

16. Salm S, Hower K, Neumann S, et al. Validation of the German version of the transsexual voice questionnaire for male-to-female transsexuals. *J Voice*. 2019. In press.

17. D'haeseleer E, Behlau M, Bruneel L, et al. Factors involved in vocal fatigue: a pilot study. *Folia Phoniatr Logop*. 2016;68:112–118.

18. Titze IR. Mechanical stress in phonation. *J Voice*. 1994;8:99–105.

19. Welham NV, Maclagan MA. Vocal fatigue: current knowledge and future directions. *J Voice*. 2003;17:21–30.

20. Nanjundeswaran C, VanSwearingen J, Abbott KV. Metabolic mechanisms of vocal fatigue. *J Voice*. 2017;31. 378.e1−378.e11.

21. Vilkman E. Voice problems at work: a challenge for occupational safety and health arrangement. *Folia phoniatr et Logoped*. 2000;52:120–125.

22. Vilkman E. Occupational safety and health aspects of voice and speech professions. *Folia Phoniatr Logop*. 2004;56:220–253.

23. Szabo Portela A, Granqvist S, Ternström S, et al. Vocal behavior in environmental noise: comparisons between work and leisure conditions in women with work-related voice disorders and matched controls. *J Voice*. 2018;32. 126.e23−126.e38.

24. Eustace CS, Stemple J, Lee L. Objective measures of voice production in patients complaining of laryngeal fatigue. *J Voice*. 1996;10:146–154.

25. Chang A, Karnell MP. Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: a study of vocal fatigue. *J Voice*. 2004;18:454–466.

26. 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.

27. Boucher VJ. Acoustic correlates of fatigue in laryngeal muscles: findings for a criterion-based prevention of acquired voice pathologies. *J Speech Lang Hear Res*. 2008;51:1161–1170.

28. 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.

29. Buekers R. Are voice endurance tests able to assess vocal fatigue? *Clin Otolaryngol Allied Sci*. 1998;23:533–538.

30. Hanschmann H, Gaipl C, Berger R. Preliminary results of a computer-assisted vocal load test with 10-min test duration. *Eur Arch Otorhinolaryngol*. 2011;268:309–313.

31. Nanjundeswaran C, Jacobson BH, Gartner-Schmidt J, Verdolini Abbott K. Vocal Fatigue Index (VFI): development and validation. *J Voice*. 2015;29:433–440.

32. Naderifar E, Moradi N, Farzadi F, et al. Cross-cultural adaptation and validation of the Vocal Fatigue Index into Persian. *J Voice*. 2019. In press.

33. Athira UR, Devadas U. Adaptation and validation of Vocal Fatigue Index (VFI) to Malayalam language. *J Voice*. 2019. In press.

34. Gardini A, Poehlke T, Reimer J, et al. Cultural and linguistic validation of the questionnaire ODAS (EADO) used to determine the adequacy of the daily dose of methadone as part of a maintenance programme for the treatment of opiate dependence. *Suchttherapie*. 2010;11:138–140.

35. Harkness J. *Conducting Cross-National and Cross-Cultural Surveys: Papers from the 2005 Meeting of the International Workshop on Comparative Survey Design and Implementation (CSDI)*. Manheim: ZUMA; 2006.

36. Willis G, Zahnd E. Questionnaire design from a cross-cultural perspective: an empirical investigation of Koreans and non-Koreans. *J Health Care Poor Underserved*. 2007;18(4 suppl):197–217.

37. 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.

38. Van Houtte E, Van Lierde K, D'Haeseleer E, et al. The prevalence of laryngeal pathology in a treatment-seeking population with dysphonia. *Laryngoscope*. 2010;120:306–312.

39. Mozzanica F, Ginocchio D, Barillari R, et al. Prevalence and voice characteristics of laryngeal pathology in an Italian voice therapy-seeking population. *J Voice*. 2016;30. 774.e13−774.e21.

40. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *IJME*. 2011;2:53–55.

41. Everitt BS. *The Cambridge Dictionary of Statistics*. 2nd ed. New York: Cambridge University Press; 2002.

42. Frey LR, Botan CH, Friedman PGKG. *Investigating Communication: An Introduction to Research Methods*. Englewood Cliffs, NJ: Prentice-Hall; 1991.

43. Henking A, Bluhm C, Fahrmeir L. *Kreditrisikomessung. Statistische Grundlagen und Modellierung*. Berlin: Springer; 2006.

44. Wehberg S, Sauerbrei W, Schumacher M. Diagnosestudien: Wertigkeit der Sonographie bei der Differenzierung von gut- und bösartigen Brusttumoren bei Patientinnen mit klinischen Symptomen. In: Schumacher, Schulgen, eds. *Methodik klinischer Studien*. Berlin: Springer; 2007:319–340.

45. 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.

46. Paolillo NP, Pantaleo G. Development and validation of the voice fatigue handicap questionnaire (VFHQ): clinical, psychometric, and psychosocial facets. *J Voice*. 2015;29:91–100.

47. 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.

48. Gotaas C, Starr CD. Vocal fatigue among teachers. *Folia Phoniatr (Basel)*. 1993;45:120–129.

49. Roy N, Merrill RM, Thibeault S, et al. Voice disorders in teachers and the general population: effects on work performance, attendance, and future career choices. *J Speech Lang Hear Res*. 2004;47:542–551.

50. Behlau M, Zambon F, Guerrieri AC, et al. Epidemiology of voice disorders in teachers and nonteachers in Brazil: prevalence and adverse effects. *J Voice*. 2012;26. 665.e9−665.e18.

51. 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.

52. Seifpanahi S, Jalaie S, Nikoo MR, et al. Translated versions of Voice Handicap Index (VHI)-30 across languages: a systematic review. *Iran J Public Health*. 2015;44:458–469.

53. Moghtader M, Soltani M, Mehravar M, et al. The relationship between Vocal Fatigue Index and Voice Handicap Index in university professors with and without voice complaint. *J Voice*. 2019. In press.

54. Banks RE, Bottalico P, Hunter EJ. The effect of classroom capacity on vocal fatigue as quantified by the Vocal Fatigue Index. *Folia Phoniatr Logop*. 2017;69:85–93.