Validation of the Voice Handicap Index by Assessing Equivalence of European Translations

I.M. Verdonck-de Leeuw^a D.J. Kuik^b M. De Bodt^c I. Guimaraes^d E.B. Holmberg^e
T. Nawka^f C.A. Rosen^g A. Schindler^h R. Whurrⁱ V. Woisard^j

Departments of ^aOtolaryngology-Head and Neck Surgery and ^bClinical Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands; ^cDepartment of Otolaryngology and Communication Disorders, University of Antwerp, Edegem, Belgium; ^dSpeech Therapy Department, Escola Superior de Saude do Alcoitao, Lisbon, Portugal; ^eDepartment of Logopedics and Phoniatrics, Karolinska Institute, Karolinska University Hospital, Huddinge, Sweden; ^fDepartment of Phoniatrics and Pedaudiology, University of Greifswald, Greifswald, Germany; ^gDepartment of Otolaryngology-Head and Neck Surgery, University of Pittsburgh Voice Center, Pittsburgh, Pa., USA; ^hDepartment of Clinical Sciences 'L. Sacco', University of Milan, Milan, Italy; ⁱNational Hospital for Neurology and Neurosurgery, London, UK; ^jDepartment of Otorhinolaryngology, Head and Neck Surgery, Larrey University Hospital of Toulouse, Toulouse, France

Key Words

Voice Handicap Index • Measurement equivalence,
European translations • International

Abstract

Objective: We aimed to assess the equivalence of translations of the Voice Handicap Index (VHI). **Patients and Methods:** Confirmatory factor analysis was used to assess equivalence of the US version and several translations including (1) Dutch, (2) Flemish Dutch (Belgium), (3) UK English, (4) French, (5) German, (6) Italian, (7) Portuguese and (8) Swedish. VHI questionnaires were gathered from 1,281 subjects. Patients were classified into 11 voice lesion categories. Patients with incomplete response (4%) and patients within voice lesion categories with small numbers were excluded from further analyses, leaving a cohort of 1,052 patients from 8 countries. **Results:** The internal consistency of the VHI proved to be good. Confirmatory factor analysis across countries revealed that a 3-factor fixed measurement model best fitted the

data; the 3 subscales appeared to highly intercorrelated, especially in the US data. The underlying structure of the VHI was also equivalent regarding various voice lesions, but distinct groups were recognized with respect to the height of the VHI scores, indicating that various voice lesions lead to a diversity of voice problems in daily life. **Conclusion:** The US VHI and the translations appeared to be equivalent, which means that the results from studies from the various included countries can be compared.

Copyright © 2008 S. Karger AG, Basel

Introduction

Voice outcome can be measured by means of several voice evaluation methods. Imaging techniques, perceptual ratings and acoustical and aerodynamic analyses are possible outcome measures. A multidimensional voice analysis protocol is recommended for monitoring voice intervention and for research purposes, including a struc-

tured questionnaire on voice problems in daily life [1, 2]. Validated questionnaires on voice problems in daily life include the 30-item Voice Handicap Index (VHI) [3] and the shortened VHI-10 [4], a 5-item list for screening purposes [5], the 28-item Voice Activity and Participation Profile [6], the 10-item Voice-Related Quality of Life Questionnaire [7] and the Voice Outcome Survey [8]. Franic et al. [9] compared the psychometric properties of 4 voice-related quality of life instruments: the VHI, Voice Activity and Participation Profile, Voice-Related Quality of Life Questionnaire and Voice Outcome Survey. These instruments were evaluated on 11 measurement standards related to item information, versatility, practicability, breadth and depth, reliability, validity and responsiveness. The VHI met 7 of 11 criteria and the results of that study supported the use of VHI total scores for clinical application with individual patients.

The VHI is currently the most widely used questionnaire and has been translated and validated in several languages including Dutch, Flemish Dutch, French, German, Italian, Portuguese and Swedish. When these various versions of the VHI are applied in cross-national comparisons, accurate translations are needed in order to obtain valid results. The standard procedures for the translation process have been followed for these mentioned non-US-English language versions including a number of parallel translations, review of the translations leading to a consensus version, backward translations, comparison with the original American version and pilot testing of the translation. Besides literal translation, some European versions have also been adapted into a culturally relevant form. For example, 2 different versions for Dutch have been developed: 'Dutch' and 'Flemish Dutch' with minor adaptations for the Netherlands and Belgium, respectively. Therefore, it is useful to examine the equivalence of various translations and the original American VHI. The purpose of this study is assessing the equivalence of the various European and original American VHI versions. The study will also provide information regarding voice problems in the daily life of patients with various types of voice lesion.

Patients and Methods

Voice Handicap Index

The VHI is a validated questionnaire measuring voice problems in daily life [3]. It consists of 30 items with 5 response levels, scored 0–4. Summarizing the scores on the 30 items leads to a total VHI score, ranging from 0 to 120. A higher score corresponds to a higher degree of patient-based vocal handicap. Next to the

total VHI score, 3 subscales are identified with 10 items each: the physical (P items), functional (F items) and emotional (E items) subscales.

The VHI has been translated and validated into several languages, including Dutch, Flemish Dutch, French, German, Italian, Portuguese and Swedish. The present study reports on the results of a European project that was initiated at the PEVOC 5 conference in Graz. In past few years this project group has managed to build up a European database on the VHI with the purpose of comparing various translations of the VHI.

Patients

VHI questionnaires were gathered from 1,281 subjects and all data were sent to the central database manager in a de-identified fashion (age, gender and voice lesion category were provided only together with the scores on all VHI items). A unique identifier was assigned to each subject to protect the subject's identity. All VHI questionnaires were collected from patients at baseline (i.e. before laryngoscopic, surgical or medical voice treatment).

Patients with incomplete response [56 patients (4%)] were excluded from further analyses, leaving a cohort of 1,225 patients. They were classified into 11 voice lesion categories: functional dysphonia (234 patients), vocal fold nodules (131), structural voice lesions (polyps, cysts, edema; 448), vocal fold paresis (137), vocal fold cancer (47), spasmodic dysphonia (52), reflux laryngitis (102), psychogenic voice disorder (13), mutation falsetto voice (5), gender dysphonia (5) and voice problems due to Parkinson's disease (39); from 12 patients the voice disorder could not be traced. Because of small numbers and/or a specific voice category, patients with cancer (mostly from the Netherlands), spasmodic dysphonia (mostly from the UK and the Netherlands), Parkinson (all from Italy), psychogenic voice disorder, mutation falsetto voice and gender dysphonia and patients of whom the voice lesion could not be traced were left out of the analyses. The cohort used for further psychometric evaluation consisted of 1,052 patients from 8 countries. In total, the median age was 45 years (range = 12–86); there were 475 males (46%) and 570 (54%) females. An overview is given in table 1. The data in the present retrospective study originate from several national scientific investigations using the VHI in various voice pathology groups. Therefore, the pathology distribution differs between countries.

Statistical Analyses

The internal consistency of the VHI was assessed by Cronbach's α . Measurement equivalence was evaluated by means of multiple group confirmatory factor analysis. In confirmatory factor analysis, known factors are tested. In the present study, these are the total VHI, and the 3 physical, functional and emotional subscales as published by Jacobson et al. [3]. The software program LISREL (SSI Scientific Software International, Inc.) for structural equation modeling was used, both across countries and diagnostic categories. Four models were tested: (A) 1-dimensional VHI, (B) 3-dimensional VHI with an orthogonal structure, (C) 3-dimensional VHI with a fixed structure (in which fixed refers to the mean correlation values between the 3 dimensions as found in the present database), and (D) 3-dimensional VHI with a free structure. These 4 models were compared by means of their χ^2 values and associated degrees of freedom. The models were considered different if these values indicated substantial improvement of fit. Using the usual significance levels has to be consid-

Table 1. Overview of the patient cohort (n = 1,052) regarding country, voice disorder, gender and age (median, range)

Country	Voice disorder					Gender % male	Age years
	dysfunction	nodules	structural	paresis	laryngitis		
USA	13	6	28	11	11	30	45 (15–86)
Belgium	39	16	81	25	8	47	45 (15–73)
France		58				31	40 (15–72)
Germany	90	15	121	31	18	29	45 (12–76)
Italy	26	2	34	13		25	43 (17–78)
Netherlands			99	43	55	42	46 (13–85)
Portugal	30	27	52	4	8	29	43 (18–70)
Sweden	36	7	33	10	2	28	49 (29–81)

ered less informative, since fit values are extremely elevated at very high levels of degrees of freedom. As a guideline the criterion was used that the relative decrease in χ^2 (gain) should at least exceed the relative decrease in degrees of freedom (cost). Pearson's correlation coefficients were used to obtain more insight into intercorrelations between the subscales.

Results

Internal Consistency

The internal consistency of the 30 items of the VHI proved to be good, with Cronbach's α ranging from 0.91 to 0.97 for the various VHI versions. Cronbach's α of the total cohort was 0.95. Cronbach's α of the physical subscale was 0.87, of the functional subscale 0.88 and of the emotional subscale 0.89.

Confirmatory Factor Analysis

Confirmatory factor analysis was used to compare the 4 models across countries and diagnostic categories.

Regarding country, the goodness of fit statistics of the 1-factor model without any correlated error terms revealed a moderate fit to the data ($\chi^2 = 10,451$, d.f. = 3,660). A three-factor orthogonal measurement model had a significantly worse fit to the data ($\chi^2 = 11,388$, d.f. = 3,663). The 3-factor fixed structure model revealed a better fit ($\chi^2 = 9,446$, d.f. = 3,657), indicating a gain of 9.6% over a cost of 0.08%. Further fitting of the 3-factor free model revealed improvement (gain 1.5%) but at a high cost (1.1%) ($\chi^2 = 9,305$, d.f. = 3,615).

Regarding voice category, similar results were found: the goodness of fit statistics of the 1-factor model without any correlated error terms revealed a moderate fit to the data ($\chi^2 = 7,458$, d.f. = 2,265). A 3-factor orthogonal mea-

Table 2. Overview of intercorrelations between the physical (P), functional (F) and emotional (E) subscales regarding the various countries

	P-F	F-E	P-E
USA	0.99	0.92	0.98
Belgium	0.80	0.91	0.77
France	0.57	0.75	0.51
Germany	0.82	0.92	0.84
Italy	0.68	0.82	0.72
Netherlands	0.66	0.87	0.62
Portugal	0.84	0.89	0.79
Sweden	0.85	0.92	0.76

surement model had a worse fit to the data ($\chi^2 = 8,295$, d.f. = 2,268). The 3-factor fixed measurement model revealed a significantly better fit ($\chi^2 = 6,373$, d.f. = 2,262), indicating a gain of 14.5% over a cost of 0.13%. Further fitting of the 3-factor free model revealed improvement (gain 1.0%) but at a high cost (1.06%) ($\chi^2 = 6,311$, d.f. = 2,238).

Intercorrelations

Substantial intercorrelations between the 3 subscales were found varying across country (table 2) and diagnostic category (table 3).

To provide information on how patients from various countries and patients suffering from various voice disorders encounter voice problems in daily life, a bar chart of the VHI subscale scores regarding country is shown in figure 1 and regarding voice lesion category in figure 2.

Regarding country, all 3 subscales were very highly intercorrelated (>0.90) in the USA. The functional and

Fig. 1. Bar chart of the mean total VHI score and the subscales emotional, functional and physical regarding the various countries.

Fig. 2. Bar chart of the mean total VHI score and the emotional, functional, and physical subscales regarding all voice disorder categories, including those that were excluded from the statistical analyses (cancer, spasmodic dysphonia, Parkinson, psychogenic, mutation falsetto and gender dysphonia).

Table 3. Overview of intercorrelations between the physical (P), functional (F) and emotional (E) subscales regarding the various voice lesion categories

	P-F	F-E	P-E
Structural	0.77	0.91	0.74
Dysfunction	0.83	0.95	0.81
Nodules	0.64	0.87	0.60
Laryngitis	0.73	0.89	0.70
Paresis	0.82	0.84	0.69

emotional subscales were highly intercorrelated in Flanders, Sweden and Germany. Figure 1 shows a bar chart of the VHI subscale scores regarding country. With the exception of the USA, the physical subscale is scored highest in all other countries. In most countries, little difference was found between the functional and emotional subscale scores.

Regarding voice disorder category, no high (>0.90) intercorrelations were found between the physical and

functional subscales, nor between the physical and emotional ones. The functional and emotional subscales were highly intercorrelated in patients with structural voice disorder and vocal dysfunction. Figure 2 shows a bar chart of the VHI subscale scores regarding voice lesion category (all categories are shown here, including those that were excluded from statistical analyses due to small numbers). Again, the physical VHI subscale is scored highest, followed by the functional and emotional subscales, except in gender dysphonia. Patients suffering from vocal fold paresis, spasmodic dysphonia, psychogenic voice disorder, mutation falsetto voice and gender dysphonia have higher VHI scores compared to patients suffering from vocal dysfunction, vocal nodules, structural lesions, laryngitis or laryngeal cancer.

Conclusion

The primary aim of this study was to assess the equivalence of the various translated versions of the VHI. Based on the confirmatory factor analyses of the total

VHI and the 3 physical, functional and emotional subscales, it can be concluded that a 3-dimensional model provides a better fit of the data than a 1-dimensional model. However, all 3 subscales were highly intercorrelated, especially so in the US data. This may lead to the conclusion that the total VHI score can also be used to assess the impact of voice problems in daily life. To obtain insight into how the VHI is used in research projects, we performed a global inspection of peer-reviewed papers as retrieved through Pubmed in the first 4 months of 2006 (search term: Voice Handicap Index). This inspection also revealed a 2-fold application of the VHI: among 13 studies, 4 reported results on the total VHI score, and 9 studies reported results on the subscales. Cohen et al. [10] published a meta-analysis on studies up to 2004 and reported on the total VHI score; it is not clear whether the subscale data were not available or not included in the meta-analysis. Bogaardt et al. [11] proposed a 2-dimensional VHI structure based on Rasch analysis; however, in Rasch analysis the VHI 5-level response scale is reduced to a dichotomous scale with 2 response levels, which leads to a loss of information. In the first 4 months of 2006, 1 study reported on the VHI-10. The VHI-10 was developed based on the idea that a shorter questionnaire would be more feasible in clinical practice [4]. That study revealed a 1-dimensional structure underlying the VHI, justifying the application of a short 1-dimensional VHI-10. However, the study was performed on data of the same institute in Pittsburgh that contributed data to the present study. The question arises whether these somewhat different findings regarding the subscale intercorrelations between the USA and Europe are due to cultural factors or differences in patient mix. From a cultural point of view, an example is the item on loss of income due to a voice problem. In the USA this appears to be an important issue (this item was also included in the VHI-10) but not in some of the European countries. For instance in the Netherlands, this item is not an issue for over 90% of the patients. This difference may be explained by the good social-economic security system in the Netherlands. Future research involving more databases from the USA may provide insight. Another explanation for the minor differences regarding subscale intercorrelations between the countries may be the different group proportions and characteristics of the patient cohorts provided by the various countries. Although a large database including more than 1,000 patients was used in the present study, the proportion of lesion categories was not optimally balanced. For instance, the data of France consisted of patients with vocal fold nodules only. Also, in the

present study data were not corrected for other aspects that may influence voice-related problems in daily life such as age, gender, duration of symptoms or occupation.

A secondary aim of this study was to provide information regarding the extent of voice problems in the daily life of patients with various voice disorders. The physical VHI subscale was most often scored highest, followed by the functional and emotional subscales. Patients suffering from spasmodic dysphonia, vocal fold paresis, psychogenic voice disorder, mutation falsetto voice and gender dysphonia appeared to have higher VHI scores compared to patients suffering from vocal dysfunction, vocal nodules, structural lesions, laryngitis, laryngeal cancer or Parkinson. In a meta-analysis, Cohen et al. [10] compared the VHI scores of 2 groups of patients: patients with neurologically based diseases (such as spasmodic dysphonia, vocal fold paralysis and other neurological voice disorders as Parkinson), and patients with inflammatory or traumatically induced voice disorders (like polyps, cysts, edema, nodules or reflux laryngitis). Their meta-analysis revealed that patients with inflammatory or traumatically induced voice disorders have significantly lower VHI scores compared to the neurological group, especially the patients with spasmodic dysphonia, which is consistent with the results shown in the present paper. Their finding that patients with Parkinson's disease encounter the severest voice problems in daily life was not confirmed.

The US VHI and the various non-US-English European translations appeared to be equivalent, which means that the results from studies from the various included languages are comparable. The underlying structure of the VHI was also equivalent regarding various voice disorders, but distinct groups were recognized with respect to the height of the VHI scores, indicating that the various voice disorders lead to a diversity of voice problems in daily life.

Acknowledgments

Chad Gundy is highly acknowledged for his contribution to the statistical analyses. Thanks to Hans Mahieu and Rico Rinkel for referring Dutch patients, to all speech and language therapists and ENT surgeons in Portugal who provided the Portuguese data, to Severine Bodin in France, to Ute Gonnermann for preparing the German data for analysis, and to the speech and language pathologists in Sweden who provided the Swedish VHI data. In Italy, thanks to Elena Favero and Cinzia Bachman for their help in collecting data. In the USA, thanks to research associate Jaime Osborne.

References

- 1 Behrman A: Common practices of voice therapists in the evaluation of patients. *J Voice* 2005;19:454–469.
- 2 Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, Van De Heyning P, Remacle M, Woisard V, Committee on Phoniatrics of the European Laryngological Society (ELS): 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 Jacobson BH, Johnson A, Grywalski C, Silbergleit G, Jacobson G, Benninger MS, Newman CW: The voice handicap index (VHI): development and validation. *Am J Speech Lang Pathol* 1997;6:66–70.
- 4 Rosen CA, Lee AS, Osborne J, Zullo T, Murray T: Development and validation of the voice handicap index-10. *Laryngoscope* 2004;114:1549–1556.
- 5 Van Gogh CD, Verdonck-de Leeuw IM, Boon-Kamma BA, Langendijk JA, Kuik DJ, Mahieu HF: A screening questionnaire for voice problems after treatment of early glottic cancer. *Int J Radiat Oncol Biol Phys* 2005;62:700–705.
- 6 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.
- 7 Hogikyan ND, Sethuraman G: Validation of an instrument to measure voice-related quality of life (V-RQOL). *J Voice* 1999;13:557–569.
- 8 Gliklich RE, Glovsky RM, Montgomery WM: Validation of a voice outcome survey for unilateral vocal cord paralysis. *Otolaryngol Head Neck Surg* 1999;120:153–158.
- 9 Franic DM, Bramlett RE, Bothe AC: Psychometric evaluation of disease-specific quality of life instruments in voice disorders. *J Voice* 2005;19:300–315.
- 10 Cohen SM, Dupont WD, Courey MS: Quality-of-life impact of non-neoplastic voice disorders: a meta-analysis. *Ann Otol Rhinol Laryngol* 2006;115:128–134.
- 11 Bogaardt HC, Hakkesteegt MM, Grolman W, Lindeboom R: Validation of the voice handicap index using Rasch analysis. *J Voice* 2007;21:337–344.