Applying the Voice Handicap Index (VHI) to Dysphonic and Nondysphonic Hebrew Speakers

*Ofer Amir, *Odelia Ashkenazi, *Tali Leibovitzh, *Orit Michael,
*Yael Tavor, and *†Michael Wolf

*†*Tel-Aviv, Israel*

Summary: Objective: To evaluate a translated version of the Voice Handicap Index (VHI) as a diagnostic tool for people with and without a laryngeal pathology, among Hebrew speakers. Study Design: Parallel group design. Methods: The VHI was translated and adapted to Hebrew. The translated version was, then, administered to a group of 182 patients with various laryngeal pathologies and a control group of 171 people with no laryngeal pathology. Based on the participants' responses to the VHI, statistical analyses were, initially, performed to assess validity and reliability, and then to evaluate group differences between the pathological and control groups and among the different pathological groups included in the study. Results: Statistical analyses showed high reliability values of the Hebrew version of the VHI (overall Cronbach's alpha $r = 0.976$). Participants' scores were not affected by their age ($P = 0.156$) or gender ($P = 0.261$). The participants in the control group obtained significantly lower scores on the overall VHI score, as well as on all three subscale scores, in comparison with the pathological group ($P < 0.001$). In addition, within the pathological group, patients with neurogenic pathologies received higher scores than all other pathological groups, whereas patients with laryngeal inflammation received lower scores than all other pathological groups ($P < 0.05$). Conclusion: The VHI is a powerful tool for quantifying patients' perceptions of their voice handicaps, and it maintained its power across translation. The VHI was shown to be valuable for the assessment of speakers with, as well as without laryngeal pathologies.

Key Words: Voice—VHI—Subjective evaluation—Questionnaire—Hebrew.

INTRODUCTION

Evaluation of voice disorders can be performed using multiple approaches, which include laryngoscopic techniques, such as stroboscopy,[1] electromyography,[2] imaging techniques,[3] aerodynamic measurements,[4] acoustic analyses,[5] subjective listener evaluations,[6] and measures of functional disability that are self-evaluated by the

speaker.[7] Although the importance and clinical value of the objective and instrumental measures have been established in numerous studies, there is a growing body of research and clinical work that suggests that these measures fail to assess the level of disability experienced by the speaker as a function of a voice disorder.[8] It was also suggested that the subjective evaluation of a voice problem, made by the patient, is dependent on a wide range of parameters, such as the individual's overall daily function, occupation, social interactions, and psychological set.[9]

Several measures have been developed as instruments for self-evaluation of voice problems. These measures include, for example, the Voice Symptom Scale (VoiSS),[10] the Voice Related Quality of Life Measure (V-RQOL),[11] the Vocal Performance Questionnaire (VPQ),[12] and the Voice Handicap Index (VHI).[7] Of these questionnaires, the VHI was widely accepted and used for research as well as for clinical application. Furthermore, the VHI was acknowledged in 2002 by the Agency of Healthcare Research and Quality as a reliable and valid diagnostic tool.[13] Since then, the VHI has been translated and adapted to German,[14] Chinese,[15] Portuguese,[16] and Polish.[17] These studies and others have demonstrated the VHI to reliably quantify the subjective perception of handicap associated with the voice problem experienced by the speaker. Although most studies used the VHI on patients with various laryngeal pathologies, only two studies have compared patients with pathologies with controls.[9,16] The primary purpose of this study was therefore to assess the validity of the Hebrew version of the VHI. In addition, the secondary purpose of this study was to evaluate quantifiable differences in VHI scores among speakers with different laryngeal pathologies and speakers without voice complaints.

METHODS

Translation and adaptation

The original English version of the VHI[7] was translated from English to Hebrew using a similar procedure to that performed previously in other languages.[14,16] To that end, four native speakers of Hebrew, who are also highly proficient in written and spoken English, performed the English-to-Hebrew translation. Because the VHI is intended for use by lay persons, the use of professional terminology was not considered to be desirable. Therefore, it was decided to use translators who are not professionally knowledgeable in the field of speech and voice. This resulted in four different Hebrew working-versions of the VHI. The four Hebrew versions were, then, translated back to English by four native speakers of English, who are also lay persons, highly proficient in written and spoken Hebrew. These four retranslated English versions were compared with the original VHI, individually for each item of the questionnaire. Following this procedure, a final Hebrew version was obtained, which was assembled by the items that translated most accurately from English to Hebrew and then back to English. The final Hebrew version (see Appendix A) was then presented, along with the original version, to four English–Hebrew bilingual judges. These judges confirmed that the final Hebrew version is comparable with the original English version.

Subjects

After obtaining the approval of our institutional review board, and an informed consent of all participants, a total of 353 speakers (144 men and 209 women), with a minimum age of 16 years, were included in this study. Of these participants, 182 were included in the pathological groups and 171 were included in the control group. The control group consisted of nondysphonic persons who were recruited in the Tel-Aviv area. The VHI results of these persons were included in the study only if the subjects denied having any history of voice problems, complaints of voice function, or history of speech or voice therapy. In contrast, all patients included in any of the pathological groups had reported a voice complaint and were examined by laryngologists in different medical centers in the Tel-Aviv area.

The participants in the pathological groups were assigned to the six pathological groups, based on the diagnoses obtained during the laryngeal examinations. The "mass lesions" group consisted of patients who were found to have uni- or bilateral nodules, polyps, cysts, or granulomas. The "inflammation" group consisted of patients who were

found to have Reinke's edema, chronic laryngitis, swelling of the laryngeal mucosa, and gastroesophageal reflux (GERD). The ''mucosa irregularity'' group consisted of patients who were found to have disturbances or asymmetric mucosal wave patterns, *sulcus vocalis*, and superficial vocal fold scar. The ''neurogenic'' group consisted of patients who were found to have uni- or bilateral vocal fold paralysis or paresis and spasmodic dysphonia. The ''functional'' group consisted of patients who were found to have a normal structural larynx but hyper- or hypofunctional laryngeal mobility patterns. Finally, the ''other'' group consisted of patients who had different laryngeal pathologies or abnormalities, which could not be assigned to any of the other groups. Table 1 presents gender, age, height, and weight information for each of the experimental groups.

RESULTS

Group means and standard deviations of the VHI subscales and total score are presented in Table 2.

The data presented in Table 2 show a consistent trend of lower values of the total and subscales scores for the control group, in comparison with all pathological groups. In addition, the severity ratings of the six pathological groups maintained identical hierarchy for the three subscales as well as for the total score.

Separate univariate analyses of variance for the total score and for the functional, physical, and emotional subscales revealed a significant main effect for group [($F_{6, 347} = 78.99$, $P < 0.001$), ($F_{6, 347} = 41.43$, $P < 0.001$), ($F_{6, 347} = 109.23$, $P < 0.001$), and ($F_{6, 347} = 54.36$, $P < 0.001$), respectively]. Post hoc analyses, using Tukey's HSD, revealed a consistent group difference between the control group and all pathological groups ($P < 0.05$) for all subscales as well as for the total score. Significant group differences were also found among different sets of experimental groups. However, these differences were not consistent across the different subscales. Detailed results of the post hoc analyses, of the three subscales as well as of the total score, are provided in Appendix B.

To assess the magnitude of the relationship among the three subscales and total scores, Pearson correlation coefficients were calculated (Table 3). Results indicated a moderate-to-strong relationship among the three subscales, ranging from $r = 0.76$ to $r = 0.86$.

The reliability of the Hebrew version was examined, first, by means of Cronbach's alpha coefficient. The resulting overall alpha coefficient for the complete VHI was $r = 0.976$. Similarly, Cronbach's alpha coefficients for the functional, physical, and emotional subscales were $r = 0.939$, $r = 0.961$, and $r = 0.953$, respectively. To evaluate parallel forms reliability, each subscale was divided into two equal parts, similar in content and number

TABLE 1. *Gender, Age (years), Height (cm), and Weight (kg) Distribution in the Experimental Groups*

Group	Gender	Age (years)	Height (cm)	Weight (kg)
Control	Male (n = 69)	39.45 ± 14.70	176.32 ± 8.34	79.40 ± 12.55
	Female (n = 102)	37.25 ± 13.55	162.73 ± 6.26	62.46 ± 14.83
Inflammation	Male (n = 13)	42.23 ± 17.04	171.42 ± 6.01	79.75 ± 16.33
	Female (n = 14)	47.00 ± 17.17	162.70 ± 7.45	59.6 ± 9.71
Other	Male (n = 16)	61.31 ± 14.34	173.46 ± 6.15	77.79 ± 10.30
	Female (n = 12)	51.83 ± 14.02	159.78 ± 4.76	61.50 ± 9.40
Functional	Male (n = 12)	37.82 ± 18.26	171.25 ± 7.31	70.83 ± 14.23
	Female (n = 16)	34.25 ± 16.58	161.00 ± 5.73	61.93 ± 10.64
Mass lesions	Male (n = 17)	41.12 ± 13.38	166.88 ± 26.45	84.38 ± 27.33
	Female (n = 50)	36.68 ± 10.93	162.50 ± 7.32	64.96 ± 18.60
Mucosa irregularity	Male (n = 3)	59.00 ± 1.00	164.00 ± 5.54	70.00 ± 17.32
	Female (n = 9)	38.33 ± 14.28	161.33 ± 4.64	65.56 ± 11.02
Neurogenic	Male (n = 15)	52.00 ± 15.20	176.77 ± 8.51	78.31 ± 12.70
	Female (n = 6)	49.50 ± 12.97	160.00 ± 9.35	76.60 ± 27.44

TABLE 2. *Group Means and Standard Deviations of the VHI Subscales and Total Score*

Group	Functional	Physical	Emotional	Total
Control	3.06 ± 2.90	3.22 ± 3.96	1.14 ± 2.34	7.55 ± 7.55
Inflammation	9.41 ± 10.39	10.16 ± 10.16	9.41 ± 10.67	35.52 ± 28.84
Other	10.93 ± 8.93	18.07 ± 8.91	11.00 ± 11.03	40.00 ± 24.94
Functional	13.41 ± 11.39	21.16 ± 11.25	15.54 ± 11.95	50.11 ± 31.36
Mass lesions	13.52 ± 10.00	25.84 ± 9.74	16.36 ± 10.49	55.72 ± 26.94
Mucosa irregularity	14.42 ± 4.83	25.92 ± 8.08	18.13 ± 6.77	58.46 ± 15.55
Neurogenic	25.19 ± 9.56	28.29 ± 8.90	22.55 ± 10.94	76.02 ± 25.69

of items. Spearman correlation coefficients were calculated between the similar split versions of the questionnaire. The resulting correlations for the functional, physical, and emotional subscales were $r = 0.889$, $r = 0.913$, and $r = 0.879$, respectively.

For validation of the questionnaire, all participants responded to two general questions, regarding their satisfaction with their voice, in addition to completing the questionnaire. The first question, "How much are you troubled by your voice?", was presented to the participants before completing the questionnaire. Responses for this question were given on a seven-point scale, where 1 was labeled "not at all" and 7 was labeled "very much." The other general question, "How satisfied are you with your voice?", was presented to the participants after the completion of the questionnaire. Responses for this question were given on a 10-point scale, where 1 was labeled "completely dissatisfied" and 10 was labeled "highly satisfied." The validity of the questionnaire was assessed by computing the Spearman correlation coefficients between the two general questions and the VHI-Heb total score. Correlation between the total score and the first general question was $r = 0.799$. Correlation between the total score and the second general question was $r = -0.769$. In addition, correlation

between the two general questions was $r = -0.766$.

Finally, responses on the Hebrew version were found not to be affected by the respondent's gender and age. Specifically, no main effect was found for Gender ($F_{1,332} = 1.387$, $P = 0.261$); and no significant correlation was found between the VHI total score and the respondent's age ($r = 0.077$, $P = 0.156$).

DISCUSSION

This study presents an adapted Hebrew version of the VHI. Before evaluating the differences among pathological groups and between the pathological and control groups, it was essential to evaluate the reliability and validity of this version of the questionnaire. Results indicated that the adapted Hebrew version was as reliable as the original version.[7] In addition, results of the questionnaire showed high correlation with the two general self-evaluation questions. These two questions were constructed such that in one question, higher response values represented high satisfaction, whereas the opposite was true for the other question. Furthermore, although one question used a 7-point scale, the other used a 10-point scale. Results indicated that despite these methodological differences, similar correlation coefficient values were obtained between these scales and the VHI total score.

The authors of the original VHI reported that they had to eliminate several items, in the process of constructing the final version of the questionnaire, because responses to these items differed between men and women.[7] We were therefore interested to verify that the current version was also unaffected by gender. This issue is of special

TABLE 3. *Correlation Matrix for Total Score and for the Three Subscale Scores for the Voice Handicap Index*

Subscale	Functional	Physical	Emotional	Total
Functional	–	0.757	0.855	0.919
Physical	–	–	0.829	0.933
Emotional	–	–	–	0.953

interest for languages that use morphology linguistic markers for gender (eg, Hebrew). Results obtained in this study revealed no gender differences in responding to the adapted Hebrew version, which is similar to the results obtained in English.

Although most studies in the field compared various groups of patients with different laryngeal pathologies, our study included a large group of controls. Data show that people with laryngeal pathologies responded to the VHI with significantly higher scores than those obtained in the control group. It was also noted that although overall scores within the pathological groups ranged between 0 and 120, the control group's overall scores ranged between 0 and 37. This result, which is based on a large group of participants, supports that of the only other two studies[9,16] that compared VHI scores between dysphonic and control participants. Thus, our data confirm that the VHI effectively differentiates pathological from healthy subjects. It is, therefore, concluded that high scores on the VHI are typical only of dysphonic patients, whereas low scores on the VHI can be obtained in dysphonic as well as in nondysphonic participants.

The VHI is not intended to distinguish between different pathological groups, because it is aimed at quantifying *subjective perception* of the voice problem, made by the patient.[7] However, our data present intriguing findings about the relationship between the laryngeal pathology and subjective self-evaluation of vocal handicap. Within the six pathological groups included in this study, a consistent hierarchy of subjective severity of handicap was obtained, using the overall score as well as all three subscale scores. For example, the neurogenic disorders group always received the highest scores, and the inflammation group always received the lowest scores. This result provides support for the need to view voice disorders as a multidimensional phenomenon. Namely, although the VHI is intended for measuring subjective self-evaluation of vocal function, it is also strongly influenced by the organic nature of the disorder. Thus, although the VHI differentiates well between pathological and healthy subjects, it should be interpreted with caution when administered to different pathological groups. Furthermore, this shows that the VHI is useful in complementing other available diagnostic tools of vocal function. However, implementing this tool independently might present misleading results.

CONCLUSION

This study demonstrated that the Hebrew version of the VHI is valid and reliable for quantifying the perception of voice handicap of patients with laryngeal pathologies. Furthermore, although most previous studies have limited the use of the VHI to patients with different pathologies, this study has shown that this tool can be used both with people who have or do not have laryngeal pathology or voice complaints. Combining the results obtained from this tool with other well-established diagnostic tools such as stroboscopy, aerodynamic evaluation, and acoustic analyses provides a multidimensional perspective on voice disorders and enhances our capabilities to assess vocal function and voice production.

REFERENCES

1. Woo P. Quantification of videostrobolaryngoscopic findings: measurements of the normal glottal cycle. *Laryngoscope*. 1996;106:1–16.
2. AAEM Laryngeal Task Force. Laryngeal electromyography: an evidence-based review. *Muscle Nerve*. 2003;28:767–772.
3. deBoer JR, Pruim J, Burlage F, et al. Therapy evaluation of laryngeal carcinomas by Tyrosine-PET. *Head Neck*. 2003;25:634–644.
4. Granqvist S, Hertegard S, Larsson H, Sundberg J. Simultaneous analysis of vocal fold vibration and transglottal airflow: exploring a new experimental setup. *J Voice*. 2003;17:319–330.
5. Baken RJ. *Clinical Measurement of Speech and Voice*. Needham Heights, MA: Allyn and Bacon; 1987.

6. Yamaguchi H, Shrivastav R, Andrews ML, Niimi S. A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale. *Folia Phoniatr Logop*. 2003;55:147–157.

7. Jacobson BH, Johnson A, Grywalski C, et al. The voice handicap index (VHI): development and validation. *Am J Speech Lang Pathol*. 1997;6:66–70.

8. Benninger MS, Atiuja AS, Gardner G, Graywalski C. Assessing outcomes for dysphonic patients. *J Voice*. 1998;12:540–550.

9. Rosen CA, Lee AS, Osborne J, Zullo T, Murry T. Development and validation of the Voice Handicap Index-10. *Laryngoscope*. 2004;114:1549–1556.

10. Deary IJ, Wilson JA, Carding PN, MacKenzie K. VoiSS: a patient-derived voice symptom scale. *J Psychosom Res.*. 2003;54:483–489.

11. Murry T, Medrado R, Hogikyan ND, Aviv JE. The relationship between ratings of voice quality and quality of life measures. *J Voice*. 2004;18:183–192.

12. Deary IJ, Webb A, Mackenzie K, Wilson JA, Carding PN. Short, self-report voice symptom scales: psychometric characteristics of the Voice Handicap Index-10 and the Vocal Performance Questionnaire. *Otolaryngol Head Neck Surg*. 2004;131:232–235.

13. Agency for Healthcare Research and Quality.Criteria for determining disability in speech-language disorders. Evidence Report/Technology Assessment. January 2002, Number 52.

14. Nawka T, Wiesmann U, Gonnermann U. Validierung des Voice Handicap Index (VHI) in der deutschen fassung. *HNO*. 2003;51:921–929.

15. Hsiung MW, Lu P, Kang BH, Wang HW. Measurement and validation of the Voice Handicap Index in voice disordered patients in Taiwan. *J Laryngol Otol*. 2003;117:478–481.

16. Guimarães I, Abberton E. An investigation of the Voice Handicap Index with speakers of Portuguese: preliminary data. *J Voice*. 2004;18:71–82.

17. Pruszewicz A, Obrebowski A, Wiskirska-Woznica B, Wojnowski W. Complex voice assessment: Polish version of the Voice Handicap Index (VHI). *Otolaryngol Pol*. 2004;58:547–549.

APPENDIX A

VHI (Hebrew Version)—VHI-*Heb*

<div dir="rtl">

הנחיית: המשפטים הבאים נאמרו על ידי אנשים רבים, כדי לתאר את קולם ואת ההשפעה שיש לקול שלהם עם חייהם. נא להקיף בעיגול את המספר המתאר באיזו תדירות המשפט נכון עבורך.

0=אף פעם   1=לעיתים רחוקות   2=לפעמים   3=לעיתים קרובות   4=תמיד

חלק H

4 3 2 1 0   1. הקול שלי מקשה על אנשים לשמוע אותי
4 3 2 1 0   2. אנשים מתקשים להבין אותי בחדר רועש
4 3 2 1 0   3. משפחתי מתקשה לשמוע אותי כשאני קורא להם מקצה הבית
4 3 2 1 0   4. אני משתמש בטלפון פחות ממה שהייתי רוצה
4 3 2 1 0   5. אני נוטה להמנע מקבוצת אנשים בגלל הקול שלי
4 3 2 1 0   6. אני מדבר פחות עם חברים, שכנים או קרובים בגלל הקול שלי
4 3 2 1 0   7. אנשים מבקשים ממני לחזור על דברי בשיחה פנים-אל-פנים
4 3 2 1 0   8. בעיות הקול שלי מגבילות את חיי האישיים והחברתיים
4 3 2 1 0   9. אני לא מרגיש שותף בשיחות בגלל הקול שלי
4 3 2 1 0   10. בעיית הקול שלי גורמת לי הפסד הכנסה

חלק P-

4 3 2 1 0   11. נגמר לי האוויר כשאני מדבר
4 3 2 1 0   12. הצליל של הקול שלי משתנה במהלך היום
4 3 2 1 0   13. אנשים שואלים, "מה קרה לקול שלך?"
4 3 2 1 0   14. הקול שלי נשמע חורק ויבש
4 3 2 1 0   15. אני מרגיש שעלי להתאמץ על מנת ליצור קול
4 3 2 1 0   16. לא ניתן לצפות מתי הקול שלי יהיה צלול
4 3 2 1 0   17. אני מנסה לשנות את הקול כדי להשמע אחרת
4 3 2 1 0   18. אני משקיע מאמץ רב בדיבור
4 3 2 1 0   19. הקול שלי גרוע יותר בערב
4 3 2 1 0   20. הקול שלי "בוגד בי" באמצע הדיבור

חלק E–

4 3 2 1 0   21. אני מתוח בזמן שאני מדבר עם אחרים בגלל הקול שלי
4 3 2 1 0   22. נראה לי שהקול שלי מפריע לאנשים
4 3 2 1 0   23. נראה לי שאנשים לא מבינים את בעית הקול שלי
4 3 2 1 0   24. בעיית הקול שלי מטרידה אותי
4 3 2 1 0   25. אני פחות פתוח בגלל בעית הקול שלי
4 3 2 1 0   26. הקול שלי גורם לי להרגיש נכה
4 3 2 1 0   27. מטריד אותי כשאנשים מבקשים ממני לחזור על דברי
4 3 2 1 0   28. אני נבוך כשאנשים מבקשים ממני לחזור על דברי
4 3 2 1 0   29. הקול שלי גורם לי להרגיש חסר יכולת
4 3 2 1 0   30. אני מתבייש בבעית הקול שלי

</div>

APPENDIX B

Results of the post hoc analyses, using Tukey's HSD with significance level set at $P = 0.05$, performed among the different experimental groups, for the three VHI subscales and total score.

A. Functional Subscale

Group	Statistical Subset		
	1	2	3
Control	3.06		
Inflammation		9.41	
Other		10.93	
Functional		13.41	
Mass lesions		13.52	
Mucosal irregularities		14.42	
Neurogenic			25.19
Within subset *P* values	1.000	0.168	1.000

B. Physical Subscale

Group	Statistical Subset			
	1	2	3	4
Control	3.33			
Inflammation		16.69		
Other		18.07		
Functional		21.16	21.16	
Mass lesions			25.84	25.84
Mucosal irregularities			25.92	25.92
Neurogenic				28.29
Within Subset *P* values	1.000	0.315	0.243	0.899

C. Emotional Subscale

Group	Statistical Subset				
	1	2	3	4	5
Control	1.16				
Inflammation		9.41			
Other		11.00	11.00		
Functional		15.54	15.54	15.54	
Mass lesions			16.36	16.36	16.36
Mucosal irregularities				18.13	18.13
Neurogenic					22.55
Within subset *P* values	1.000	0.070	0.167	0.893	0.064

D. Overall VHI-*Heb* Score

Group	Statistical Subset				
	1	2	3	4	5
Control	7.55				
Inflammation		35.52			
Other		40.00	40.00		
Functional		50.11	50.11	50.11	
Mass lesions			55.72	55.72	
Mucosal irregularities				58.46	
Neurogenic					76.02
Within subset *P* values	1.000	0.113	0.067	0.733	1.000