Validation of the Dysphonia Severity Index in the Dr. Speech Program

***HaKyung Kim, †ShaoHua Gao, ‡,§,‖ Bin Yi, ‡,§,‖ RunJie Shi, *Qin Wan, and *ZhaoMing Huang,** *‡§‖ *Shanghai, and* †*NingBo, China*

Abstract: Purpose. The Dysphonia Severity Index (DSI) is an objective multiparameter index of voice quality that measures and describes overall voice quality. Some studies have suggested that the reliability of devices for DSI measurement should be examined. We explored the feasibility of DSI measurements using the Dr. Speech (DRS) device, verified its effectiveness for clinical voice measurements and intradevice reliability, and examined the correlation between the DSI and self-evaluations of voice problems.

Methods. Seventy adult participants (including individuals with voice problems and healthy adults) underwent objective and subjective voice assessments. These data were then used to establish a DSI_{DRS} model and test the intradevice (DRS device and Praat software) reliability. The clinical validation of the DSI_{DRS} was conducted by measuring the DSI of six other participants and comparing the observed and predicted perceived voice quality as expressed by the G score (of the GRBAS scale). Moreover, the relationship between the DSI measurements and participants' self-evaluations of voice problems was investigated by analyzing the correlation between the DSI and the Voice Handicap Index (VHI).

Results. The DSI_{DRS} discriminated 80% of participants' voice quality ratings. There were strong correlations between the DSI and variables measured by the DRS device and Praat software. Furthermore, there was no significant correlation between the DSI_{DRS} and VHI.

Conclusion. The DRS device can perform DSI measurements. Objective voice measurements and perceptual voice ratings reflected different aspects of vocal function and its effects. These factors should be considered in clinical practice settings.

Key Words: Dysphonia Severity Index−Dr. Speech device−Validation−Objective assessment−Voice Handicap Index.

INTRODUCTION

The formation of voice is inseparable from the concept of phonation, and it is related to the frequency, amplitude, and regularity of vocal fold vibrations and the extent of vocal fold closure. Common methods of voice assessment include laryngoscopy, subjective evaluations by experts or patients themselves, and acoustic measurements carried out using various devices. Laryngoscopy is usually conducted by laryngologists, and requires sufficient experience and professional training. Subjective evaluations of phonation consist of perceptual evaluations by voice experts using scales such as the Grade-Roughness-Breathiness-Asthenia-Strain (GRBAS)[1] and Consensus Auditory Perceptual Evaluation of Voice scales.[2] Patient self-evaluations are typically carried out using questionnaires such as the Voice Handicap Index (VHI),[3] Voice-Related Quality of Life Measure,[4] and the Vocal Fatigue Index.[5] Human hearing is sensitive to voice quality; however, to subjectively define the pathological features of the voice is difficult because such definitions may be influenced by factors such as the professional background of the listener (rater), the type of speech material, and subjective bias. Therefore, objective measurements are important supplemental tools for assessing and describing voice quality. Voice quality measurement parameters can be frequency-related, amplitude-related, and aerodynamic in nature. However, one of the limitations of single-parameter measurements is that they only reflect one aspect of voice quality and cannot provide an overall description of vocal function. Some researchers have explored multiparameter voice measurement methods to describe voice quality as a whole. In 2000, Wutys et al[6] invented the Dysphonia Severity Index (DSI), a multidimensional objective index based on voice range profile, aerodynamic, and acoustic parameters, and used the G score (from the GRBAS scale) as the group variable. Four variables were selected using a stepwise regression method, and the weighted combination of these four variables was defined as the DSI, where $DSI = 0.13 \times MPT\ (seconds) + 0.0053 \times HF0\ (Hz) - 0.26 \times LI\ (dB) - 1.18 \times Jitter\ (\%) + 12.4$. Subsequent

Accepted for publication August 13, 2019.

Funding: This research has been supported by Peak Discipline Construction Project of Education at East China Normal University, by the Ministry of Education Humanities and Social Sciences Fund Project in China (grant number 17YJEZH001) and by the 2016 Higher Education Teaching Reform Project of Zhejiang Education Department (grant number jg20160375).

Conflict of Interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

From the *Department of Education and Rehabilitation, Faculty of Education, East China Normal University, Shanghai, China; †NingBo College of Health Sciences, NingBo, China; ‡Department of Otorhinolaryngology-Head and Neck Surgery, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China; §Laboratory of Auditory Neuroscience, Ear Institute, Shanghai JiaoTong University School of Medicine, Shanghai, China; and the ‖Shanghai Key Laboratory of Translational Medicine on Ear and Nose Diseases, Shanghai, China.

Address correspondence and reprint requests to RunJie Shi, Department of Otorhinolaryngology-Head and Neck Surgery, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, 639, ZhiZaoJu Road, Shanghai 200011, China.

Address correspondence and reprint requests to HaKyung Kim, Department of Education and Rehabilitation, Faculty of Education, East China Normal University, 605, Tianjiabing, North Zhongshan Road, Shanghai 200062, China.

E-mail addresses: shirunjie9h@163.com hgjin@spe.ecnu.edu.cn

Journal of Voice, Vol. 33, No. 6, pp. 948.e23−948.e29

0892-1997

https://doi.org/10.1016/j.jvoice.2019.08.011

research has demonstrated that this index can be applied in the practice of voice assessment and monitoring.[6–8]

The devices used for DSI measurements vary among studies. In the original study,[6] the measurement was completed using the Multidimensional Voice Program (MDVP) and the Voice Range Profile (VRP) program (Kay PENTAX Corp). Other devices used for DSI measurements included the Ling WAVES system (WEVOSYS, Forchheim, Germany),[9] the DiVAS system (XION, Berlin, Germany),[10] TF32 computer software,[11] and Praat (Institute for Phonetics Sciences, University of Amsterdam, the Netherlands).[8] Since the systems and algorithms likely differ across different devices, the intradevice reliability of DSI measurements should be examined. In 2012, Aichinger et al[9] compared the DSI measured by LingWAVES (WEVOSYS, Forchheim, Germany) and DiVAS (XION), and found that the DSI measured by these two devices were significantly different from each other; the obtained DSI values were higher than the values in the original study. The authors concluded it was because of hardware (eg, precision of sound meter) and software (eg, algorithm of VRP) differences between the different devices. In 2017, Maryn et al[8] investigated the feasibility of Praat software for measuring DSI and compared its DSI values with the DSI values measured by the VRP and MDVP programs (Kay PENTAX Corp). They found that the correlation of the DSI and these two devices was 0.851, indicating that the intradevice consistency of the DSI measurements was relatively high.

The Dr. Speech (DRS) device is one of the most commonly used instruments for speech assessment and therapy in mainland China. It has adopted technology for digital signal processing, can record and analyse audio signals in real-time, and provides visual and auditory feedback while recording and analyzing sounds. However, this device can only complete single-parameter measurements and has no function for DSI measurements. Still, in mainland China, the mainstream methods for voice assessment are laryngoscopy and experts' subjective evaluations. Clinicians rarely use objective parameters to describe and compare patients' voices. The scarcity of specific voice clinics could explain this situation in mainland China. Patients with voice problems generally present to otolaryngology clinics, where a doctor generally diagnoses each patient's voice disorder based on laryngoscopy and perceptual evaluations, rather than the results of acoustic measures. Therefore, even though voice measurement devices are common in mainland China, they are usually used for scientific research of voice instead of clinical diagnosis. Given this situation, the DSI has the advantage of being able to objectively reflect overall voice quality. Because these measures can be obtained relatively easily and efficiently, it can be an important supplement for voice description and monitoring; thus, the implementation of DSI measurement using a DRS device would be helpful for both clinical voice diagnosis and research.

This study aimed to explore the feasibility of the DRS device for measuring DSI. We established a DSI_{DRS} model and examined its effectiveness by comparing it with the GRBAS perceptual severity rating. We also examined the intradevice (DRS device and Praat software) reliability of DSI measurements. Finally, in order to evaluate the relationship between DSI and self-evaluation, we also analyzed the correlation between the DSI and VHI.

MATERIALS AND METHODS

Research participants

All data that were used for the establishment of the DSI_{DRS} model and correlation analysis between objective and subjective voice evaluations were gathered from patients from the outpatient clinic of the Department of Otorhinolaryngology in the affiliated Ninth People's Hospital of Shanghai Jiaotong University and healthy individuals (graded G0 on the GRBAS scale by speech pathologists). There were 33 males and 37 females; their age ranged from 18 to 74 years (mean age, 41.41 years; standard deviation, 15.93 years). Furthermore, these participants had not received voice therapy previously. Additionally, the data of six additional participants (four females and two males with a mean age of 40.5 years and a standard deviation of 13.9 years) whose voice quality had not been grouped were also used for validating the DSI_{DRS} model. All individuals participated in this study voluntarily, and before the voice assessment procedure, the participants were informed of the entire procedure and purpose and signed the informed consent form willingly. This study was performed with the informed consent of all participants, and the Institutional Ethics Committee approved the study protocol (approval number HR183-2017 & SH9H-2018-T12-1).

Sound recording and voice assessment

The DRS device (Tiger Rehabilitation Medical Technology Co., Ltd., Shanghai) and Praat software (Institute for Phonetics Sciences, University of Amsterdam, the Netherlands) were used simultaneously to measure voice parameters. The Real-Time Speech Analysis and Voice Assessment and Analysis programs in the DRS device were used to record and analyse voice samples. The device was connected with a low-pass filter with a sampling frequency of 44,100 Hz, and 16 bits, and the software was downloaded onto an HP envy13-AD109TU laptop, which was connected to the same type of low-pass filter to maintain the same sampling frequency as the DRS device. Two Shure-SV100C dynamic microphones (Shure Incorporate, Evanston, IL) were connected to the DRS and Praat devices and were used for the recordings. The microphones were placed at the same horizontal position with a distance of about 10 cm, with a horizontal tilt of 30°−45° and 10−15 cm away from each participant's mouth during the recordings. Before recording, the participants were instructed how to phonate properly. The MPT measure was recorded three times, and the participants were asked to phonate the vowel /a/ at a comfortable and natural loudness and pitch, and the longest trial was selected for DSI calculation. The HF0 and LI measures required the participants to phonate a sustained /a/ and to

try to achieve their highest pitch and lowest loudness, the HF0 and LI were each recorded three times, and the trials that achieved the highest pitch and lowest loudness were used to the DSI calculation. Jitter was measured by recording the phonation of a sustained /a/ in a comfortable and natural pitch and loudness for about 5 seconds for three times, and the mean jitter value of the three trials was chosen for the DSI calculation. Additionally, two speech pathologists evaluated the participants' voice qualities using the GRBAS scale. The G score was used as the dependent variable for the discrimination analysis. Finally, the participants were also asked to complete the VHI form.

Sound calibration and selection of the jitter parameter

Maryn et al[8] used the Praat program to measure DSI and indicated that the sound calibration was vital for the intensity measurement because it enabled equalization of sound intensities among different recording environments. The intensity measurement in the DRS device is a relative value that is different from the value measured by a sound meter; therefore, the sound intensity measured by the DRS device should be calibrated before calculating DSI.

The sound recording and calibration were conducted in a sound-proof room with background noise intensity lower than 35 dB. The calibration process was carried out according to the following steps. First, we used Adobe Audition 3.0 software on an HP envy13-AD109TU laptop to play the speech noise (a sustained vowel /a/) at different intensities within the phonatory range. The microphones that were linked to the DRS device and the laptop running the Praat software and sonometer (AWA 6291 real-time signal analyzer, Hangzhou Aihua Instruments Co., Ltd) captured the speech noise simultaneously. The microphones and sonometer were kept at the same horizontal level, 10 cm from each other. The loudspeaker playing the speech noise was placed in front of the microphones at a distance of 10 cm. Secondly, the intensity output in the DRS device, Praat software, and the sonometer were noted on an Excel spreadsheet; for example, when the loudspeaker played the speech noise at a certain intensity, the intensity outputs that were noted in the Excel spreadsheet were 63 dB for the DRS device, 61 dB for the Praat software, and 69 dB for the SPL in the sonometer. Thirdly, we used a linear regression analysis method with the Excel spreadsheet data to convert the measured intensities from the two different devices (DRS device and Praat software) to the actual sound intensity. The final intensity calibration formula for the DRS device was IE = 1.1388 × IM, and the calibration formula for the Praat software was IE = 1.1295 × IM.

Since there are four algorithms for the parameter jitter (%), the optimized jitter should be selected before the DSI calculation. The GRBAS scale is correlated closely with some acoustic parameters such as jitter, shimmer, and HNR.[12,13] Because of the high reliability and validation of the GRBAS scale in voice quality assessments, GRBAS is usually used as the basis of a multidimensional acoustic analysis.[6,14,15] The Spearman' rho indicated that all four algorithms of jitter generated by the DRS device were significantly correlated with the G score ($P < 0.01$), and the strongest correlation appeared between the G score and jitter PPQ (r = 0.442). The correlation between the G score and jitter %, jitter RAP, and jitter 11 pts (the other three types of jitter in DRS device) were 0.414, 0.399, and 0.390, respectively. Therefore, the value of jitter PPQ was used for the DSI calculation.

Statistical methods

We used the discrimination analysis method to establish the DSI_{DRS} model, in which the independent variables were MPT, HF0, LI, and jitter, and the dependent variable was the G score. The Pearson correlation method was used to examine the intradevice reliability of DSI_{DRS} and the relation between DSI_{DRS} and VHI. All these statistical analyses were performed with the SPSS 16.0 software.

RESULTS

Discrimination model of DSI_{DRS}

Table 1 shows the function coefficients and corresponding constants. According to the eigenvalues, the first function had the largest eigenvalue and canonical correlation value. This indicated that this function was more applicable for discrimination grouping. The canonical discriminant function coefficients indicated that there could be three discriminant functions, and according to eigenvalues, the first function was the most suitable one. Therefore, the discriminant function of DSI_{DRS} was $DSI_{DRS} = 1.829 \times \text{Jitter PPQ} - 0.005 \times \text{HF0} + 0.028 \times \text{LI} + 0.032 \times \text{MPT} - 0.831$.

Table 2 shows the Fisher's linear discriminant functions grouped by the G score. For each of the four grades of the G score, there were four corresponding Fisher discriminant functions, by which the ungrouped cases were classified into specific groups. When the four parameters of HF0, Jitter PPQ, LI, and MPT are known, the four sets of discriminant functions could be calculated. The maximum value of the four discriminant functions indicated the G-score grouping corresponding to the case.

Table 3 shows that the consistency between the original and predicted grouped cases by the discriminant functions

TABLE 1.
Canonical Discriminant Function Coefficients

	Function		
	1	2	3
HF DRS	−0.005	0.006	0.004
Jitter DRS	1.829	0.380	0.266
LI DRS	0.028	0.039	0.112
MPT	0.032	−0.101	0.101
(Constant)	−0.831	−2.346	−7.297

Unstandardized coefficients.

TABLE 2.
Classification Function Coefficients

	G Score			
	0	1	2	3
HF DRS	0.050	0.056	0.055	0.046
JITTER DRS	4.546	4.362	5.329	7.927
LI DRS	0.790	0.805	0.882	0.856
MPT	0.695	0.597	0.689	0.675
(Constant)	−30.542	−31.562	−36.099	−33.976

Fisher's linear discriminant functions.

was 54.3% (refers to the consistency between the G score obtained by GRBAS perceptual evaluation and the score obtained by the predicted grouping of discrimination analysis. This means that 54.3% of original grouped cases were correctly classified), and the 'diagonal plus one off-diagonal' classification (accept the neighboring classification of cases to be grouped, eg, the acceptable neighboring classification of G0 is Group 1, and that of G1 is Group 0 and Group 2)[6] was 85.7%.

Discrimination model of DSI_{PRAAT}

In order to obtain reliable data for the comparison of DSI values measured by the DRS device and Praat software, we used the discrimination analysis method to establish the DSI_{PRAAT} model. According to the results of the eigenvalues and the canonical discriminant function coefficients, the discriminant function of DSI_{PRAAT} is $DSI_{PRAAT} = 2.874 \times \text{Jitter PPQ} - 0.003 \times \text{HF0} + 0.036 \times \text{LI} - 0.006 \times \text{MPT} - 1.105$.

Table 4 shows that the consistency between the original grouped cases and predicted grouped cases by the discriminant functions was 47.1%. Further, the 'diagonal plus one off-diagonal' classification (accept the neighboring classification of these cases) was 81.4%.

Overall, the DSI discrimination model of the DRS device was more accurate than the Praat software.

Clinical validation of DSI_{DRS}

The clinical validation of the DSI_{DRS} model was conducted by examining the consistency between grouping results of the DSI_{DRS} discriminant function and the G score, as rated by speech pathologists. Six ungrouped individuals participated in the validation. Table 5 shows the discrimination

TABLE 3.
Classification Results

			Predicted Group Membership				
		G score	0	1	2	3	Total
Original	Count	0	8	3	3	0	14
		1	6	18	5	0	29
		2	3	5	7	1	16
		3	2	2	2	5	11
	%	0	57.1	21.4	21.4	0.0	100.0
		1	20.7	62.1	17.2	0.0	100.0
		2	18.8	31.2	43.8	6.2	100.0
		3	18.2	18.2	18.2	45.5	100.0

TABLE 4.
Classification Results

			Predicted Group Membership				
		G score	0	1	2	3	Total
Original	Count	0	9	2	3	0	14
		1	4	17	6	2	29
		2	4	10	2	0	16
		3	2	2	2	5	11
	%	0	64.3	14.3	21.4	0	100.0
		1	13.8	58.6	20.7	6.9	100.0
		2	25.0	62.5	12.5	0	100.0
		3	18.2	18.2	18.2	45.5	100.0

TABLE 5.
Comparison of Discrimination Results and G Scores of Six Ungrouped Participants

Participants	HF0	Jitter PPQ	LI	MPT	F(0)	F(1)	F(2)	F(3)	Discrimination Results	G Score
S1	188.46	0.32	0.37	29.22	0.95	−1.86	−3.56	−2.71	0	0
S2	319.57	0.16	0.18	17.09	−1.83	−2.64	−5.76	−6.35	0	0
S3	331.58	0.42	41.99	9.01	27.37	28.00	27.60	26.60	1	1
S4	432.35	0.18	45.94	20.00	42.07	42.34	42.92	40.14	2	2
S5	361.48	0.86	39.28	8.30	28.26	29.02	28.75	28.72	1	3
S6	308.07	0.27	39.28	14.036	26.91	26.90	26.63	25.46	0	0

results comparison and the G scores of these ungrouped participants.

When the values of parameters HF0, jitter PPQ, LI, and MPT are given, the values of the four discriminant functions can be obtained, and the largest function value indicates the grouping result. Compared with the G score, the consistency between the discrimination results and perceptual rating G score was 5/6 (80%), indicating that the DSI_{DRS} model can be used for voice quality discrimination.

Intra-device reliability of DSI_{DRS}

Table 6 shows the descriptive statistics results of DSI and its parameters for all 70 participants, and the results of the subjective evaluation results such as G score, VHI total score, and the subscale scores. According to the Pearson correlation analysis in Table 7, the measured DSI and all parameters between DRS and Praat were significantly correlated. Among them, there was a high correlation between HF0 measured by DRS and Praat, while the between-device correlations of jitter, LI, and DSI, measured by these two devices, was relatively lower. Since MPT is a durational measure, the values between the two devices were the same.

Correlation between DSI_{DRS} and VHI

All 70 participants underwent both the objective voice measurement (DSI) and subjective self-evaluation (VHI). The correlation between the DSI and VHI was analyzed.

Table 8 shows the correlations between the DSI value, VHI total score, and its subscale scores. There were no significant correlations between DSI, VHI, and its subscales.

Correlation between G score and VHI

The correlation between G score based on an expert's perceptual rating and VHI based on self-evaluation of participants was analyzed.

Table 9 shows that there was a significant moderate correlation between the G score and VHI total score. Among all of the subscales of VHI, the correlation between the G score and VHI functional score, and between the G score and the VHI physical score were moderate, while the correlation between the G score and VHI emotional subscale was weaker.

DISCUSSION

The applicability of the DRS device for DSI measurements

The current study aimed to examine the applicability of the DRS device for measuring DSI and the relationship between DSI and VHI. DSI, as an objective multiparameter index of voice quality, can be measured using specific acoustic measurement devices such as VRPs and perturbation measurements. In the field of voice measurements, various devices have been widely used: the original DSI version used the MDVP and CSL programs (Kay Elemetrics Corp.). After that, a variety of instruments were used to measure DSI. Researchers have recommended examining the reliability of DSI measurements among different devices. The DRS device is one of the most widely used speech assessment tools in China; however, whether it can be used for DSI measurements and its clinical applicability and intradevice reliability remain unknown. In a 2017 study that used the Praat software to measure DSI,[6] sound intensity calibration and parameter selection were necessary for the DSI measurements. Thus, in the current study, the sound level calibration and the selection of the optimum jitter algorithm were completed before we obtained the values of all parameters necessary for DSI calculation. The discrimination analysis method was used to establish the DSI_{DRS} discrimination model, and the DSI_{DRS} model was verified as a useful tool for discriminating voice quality. The results of the intradevice reliability analysis between DSI_{DRS} and DSI_{PRAAT} indicated that there was a significant positive correlation between the DSI, as measured by these two devices. This analysis produced a correlation coefficient of 0.663. This correlation is positive and moderately strong and it could be explained by the algorithm and precision variation between different devices. DSI measurement may vary among devices; however, the positive correlation between these two different devices demonstrates that the DSI measurement is relatively stable, even when calculated by different devices. Because the precision of these measures is important, future studies should continue to compare and verify DSI measurements obtained from various devices. The clinical validation of DSI_{DRS} model in the current study justifies the feasibility and effectiveness of the DRS device in DSI measurement. Also, the present study showed that the

TABLE 6.
Descriptive Statistics of DSI and Its Parameters Measured by Two Devices

	Minimum	Maximum	Mean	Standard Deviation
DSI DRS	−1.66	6.25	−0.04	1.20
DSI Praat	−1.35	7.25	0.10	1.16
HF DRS (Hz)	158.63	490.00	324.87	83.71
HF Praat (Hz)	148.09	472.55	315.96	80.69
Jitter DRS (%)	0.12	3.96	0.46	0.63
Jitter Praat (%)	0.08	2.85	0.33	0.40
LI DRS (dB SPL)	25.37	59.91	40.34	7.60
LI Praat (dB SPL)	19.31	54.05	35.47	7.65
MPT (s)	4.23	34.38	14.21	6.49
G Score	0.00	3.00	1.34	0.98
VHI Total	0.00	72.00	15.17	17.84
VHI Functional	0.00	26.00	5.21	6.60
VHI Physical	0.00	29.00	6.93	7.66
VHI Emotional	0.00	23.00	3.03	5.12

DSI discrimination model of the DRS device achieved a higher correct classification compare to the Praat software. Thus, the DRS device appears applicable for obtaining DSI measurements.

The relationship between the DSI and VHI

Voice is multidimensional, and its description should include analysis of both subjective and objective measures. The VHI evaluates an individual's voice function from emotional, physiological, and functional aspects. It is a commonly used subjective assessment tool in clinical voice assessment. Jacobson et al[3] showed that the scale was statistically robust due to its strong internal consistency and test-retest reliability.

TABLE 7.
Pearson Correlation of DSI and Its Parameters Measured by Two Devices

Pearson Correlation	Parameters	Coefficient
DRS vs Praat	HF0	0.982*
	Jitter	0.699*
	LI	0.655*
	DSI	0.663*

* Means that the *P* value is less than 0.01.

TABLE 8.
Pearson Correlation Between DSI and VHI

Pearson Correlation		Coefficient	*P*
DSI_{DRS}	VHI total	0.134	0.267
	VHI functional	0.228	0.057
	VHI physical	0.199	0.098
	VHI emotional	−0.124	0.305

Wutys et al[6] used the VHI to measure the criterion validity of the DSI and found that there was a significant negative correlation between the two indices. Hakkesteegt et al[7] studied the applicability of the DSI and VHI for evaluating the effects of voice therapy and phonosurgery and found that both DSI and VHI could reflect post-intervention changes in voice quality. Furthermore, they also demonstrated that objective and self-perceived evaluations could measure different aspects of voice and complement each other. However, some studies have indicated that there is no significant correlation between the DSI and VHI. Smits et al[16] studied the relation of vocal fold lesions and voice quality to voice handicap and psychosomatic well-being and found that—except for the significant correlations between the DSI and VHI physical subscale—there were no significant correlations between the DSI and VHI, or between the DSI and psychosomatic well-being. Therefore, the authors suggested that the objective voice measurement did not necessarily correlate with the self-perceived evaluation of voice. The present study also demonstrated that the correlation between DSI and VHI was not significant. Meanwhile, there was a strong positive correlation between the G score and VHI in the current study. These results may indicate that, even though evaluations based on experts and patients themselves vary in their evaluation content, the strong correlation between them suggests a measure

TABLE 9.
Spearman Correlation Between G Score and VHI

Spearman's Rho		Coefficient
G score	VHI total	0.467*
	VHI functional	0.420*
	VHI physical	0.467*
	VHI emotional	0.282†

* Means that the *P* value is less than 0.01.
† means that the *P* value is less than 0.05.

of consistency among different subjective evaluation methods. Compared with objective measurements, auditory perceptual and self-rating evaluations may have more flexible evaluation scales due to different assessment precision. Furthermore, objective and subjective methods of voice assessment likely reflect different aspects of vocal function, both of which should be considered in clinical practice. Moreover, the correlation results between VHI and G score indicated that, compared with the functional and physical subscales of VHI, the correlation between the emotional subscale and G score was the weakest. This phenomenon may be related to the vague concept of voice health held by the Chinese population. During the assessments, it was found that few of the participants possessed knowledge relating to voice care, and most did not report any perceived frustration or depression resulting from their voices.

CONCLUSION

The DSI_{DRS} model, based on the DRS device, can be used for voice measurements and severity discrimination. The discrimination success of DSI_{DRS} was higher than the DSI_{PRAAT} model. There was high intradevice reliability of DSI, as measured by the DRS device and Praat software. The correlation between DSI_{DRS} and VHI was not significant. The DRS device appears suitable for DSI measurements. Further, objective voice measurements and perceptual voice ratings reflect different aspects of vocal function and its effects; both methods should therefore be considered in clinical practice. There are also a few limitations to this study. Firstly, the devices used for DSI measurements included only the DRS device and Praat software. Therefore, comparisons between the DRS device and the MDVP and CSL programs (the first device used for DSI measurement) are needed. The second limitation is that due to the relatively small sample size in this study; normal and abnormal value ranges of the DSI in the Chinese population have not been confirmed. Thus, further studies could be conducted to explore the intradevice reliability of DSI measurements between the DRS device and MDVP and CSL programs. Future studies should also establish a reference DSI value range for the Chinese population.

Acknowledgments

The authors would like to thank the participants from East China Normal University and Shanghai Ninth People's Hospital for their willingness to participate in this study.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.jvoice.2019.08.011.

REFERENCES

1. Hirano M. *Psycho-Acoustic Evaluation of Voice: GRBAS Scale for Evaluating the Hoarse Voice. Clinical Examination of Voice*. Wien, New York: Springer, Verlag; 1981.
2. Gerratt BR, Kreiman J, Antonanzas-Barroso N, et al. Comparing internal and external standards in voice quality judgments. *J Speech Lang Hear Res*. 1993;36:14.
3. 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.
4. Hogikyan ND, Sethuraman G. Validation of an instrument to measure Voice-Related Quality of Life (V-RQOL). *J Voice*. 2000;13:557–569.
5. Nanjundeswaran C, Jacobson BH, Gartner-Schmidt J, et al. Vocal Fatigue Index (VFI): development and validation. *J Voice*. 2015;29:433–440.
6. Wuyts FL, Bodt MSD, Molenberghs G, et al. The Dysphonia Severity Index: an objective measure of vocal quality based on a multiparameter approach. *J Speech Lang Hear Res*. 2000;43:796–809.
7. Hakkesteegt MM, Brocaar MP, Wieringa MH. The applicability of the dysphonia severity index and the voice handicap index in evaluating effects of voice therapy and phonosurgery. *J Voice*. 2010;24:199–205.
8. Maryn Y, Morsomme D, De Bodt M. Measuring the Dysphonia Severity Index (DSI) in the Program Praat. *J Voice*. 2017;31:644.e29–644.e40.
9. Aichinger P, Feichter F, Aichstill B, et al. Inter-device reliability of DSI measurement. *Logoped Phoniatr Vocol*. 2012;37:167–173.
10. Zhou Z, Ge P, Liu Q, et al. Observation of the dysphonia severity index in evaluating curative effect of vocal cord polyp surgery. *Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi*. 2015;50:673.
11. Awan SN, Ensslen AJ. A comparison of trained and untrained vocalists on the Dysphonia Severity Index. *J Voice*. 2010;24:661–666.
12. Freitas SV, Pestana PM, Almeida V, et al. Integrating voice evaluation: correlation between acoustic and audio-perceptual measures. *J Voice*. 2015;29:390.e1–390.e7.
13. Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *J Voice*. 2004;18:299–304.
14. Takayanagi T, Akao N, Tsukidate S, et al. Acoustic parameters for classification of breathiness in continuous speech according to the GRBAS Scale. *J Voice*. 2014;28:653.e9–653.e17.
15. Wang Z, Yu P, Yan N, et al. Automatic assessment of pathological voice quality using multidimensional acoustic analysis based on the GRBAS Scale. *J Signal Process Syst*. 2016;82:241–251.
16. Smits R, Marres H, De Jong F. The relation of vocal fold lesions and voice quality to voice handicap and psychosomatic well-being. *J Voice*. 2012;26:466–470.