AJSLP

Research Article

Self-Ratings of Vocal Status in Daily Life: Reliability and Validity for Patients With Vocal Hyperfunction and a Normative Group

Jarrad H. Van Stan,[a,b,c] Marc Maffei,[a,b] Maria Lúcia Vaz Masson,[a,c,d] Daryush D. Mehta,[a,b,c] James A. Burns,[a,c] and Robert E. Hillman[a,b,c]

Purpose: The aim of this study was to establish reliability and validity for self-ratings of vocal status obtained during the daily activities of patients with vocal hyperfunction (VH) and matched controls.
Method: Eight-four patients with VH and 74 participants with normal voices answered 3 vocal status questions—difficulty producing soft, high-pitched phonation (D-SHP); discomfort; and fatigue—on an ambulatory voice monitor at the beginning, 5-hr intervals, and the end of each day (7 total days). Two subsets of the patient group answered the questions during a 2nd week after voice therapy (29 patients) or laryngeal surgery (16 patients).
Results: High reliability resulted for patients (Cronbach's $\alpha = .88$) and controls ($\alpha = .95$). Patients reported higher D-SHP, discomfort, and fatigue (Cohen's d = 1.62–1.92) compared with controls. Patients posttherapy and postsurgery reported significantly improved self-ratings of vocal status relative to their pretreatment ratings (d = 0.70–1.13). Within-subject changes in self-ratings greater than 20 points were considered clinically meaningful.
Conclusions: Ratings of D-SHP, discomfort, and fatigue have adequate reliability and validity for tracking vocal status throughout daily life in patients with VH and vocally healthy individuals. These questions could help investigate the relationship between vocal symptom variability and putative contributing factors (e.g., voice use/rest, emotions).

It is believed that diagnoses related to vocal hyperfunction (VH; true vocal fold nodules, polyps, and muscle tension dysphonia) are caused by or associated with aberrant vocal behaviors in daily life (Hillman, Holmberg, Perkell, Walsh, & Vaughan, 1989). Typical daily variation in the vocal status of patients with hyperfunction-related voice disorders has been attributed to a number of factors including amount of voice use/recovery (Nanjundeswaran, Jacobson, Gartner-Schmidt, & Abbott, 2015; Solomon, 2008), levels of emotional stress (Dietrich, Andreatta, Jiang, Joshi, & Stemple, 2012; Dietrich, Verdolini Abbott, Gartner-Schmidt, & Rosen, 2008), and presence/absence of conditions that degrade vocal function (e.g., hydration, reflux; Chung et al., 2009). One-time, in-clinic surveys of patient self-reported vocal status support the assumption that patients with VH spend significantly more time in pathological vocal states such as higher levels of vocal fatigue, discomfort in the head/neck, and lower levels of functional voice use (Hogikyan & Sethuraman, 1999; Jacobson et al., 1997; Nanjundeswaran et al., 2015). However, these surveys are prone to recall biases, which can be especially strong when the disorder symptom—such as vocal status—fluctuates over time (Streiner, Norman, & Cairney, 2014). Meaning that overall average differences between patients and subjects with normal voices may be underestimated or overestimated. Furthermore, even if the one-time ratings of overall vocal status were accurate measures, averages may portray a misleading picture of pathologically varying (i.e., nonstationary) vocal status over the course of a day or week or for specific speaking situations. Finally, any quantitative or qualitative relationships between voice use and a patient's self-rating of vocal status are limited when only using an overall one-time estimate (e.g., changes within and between objective and subjective measures are not available for investigation).

[a]Massachusetts General Hospital, Boston
[b]MGH Institute of Health Professions, Boston, MA
[c]Harvard Medical School, Boston, MA
[d]Federal University of Bahia, Brazil
Correspondence to Jarrad H. Van Stan: jvanstan@mgh.harvard.edu

Editor-in-Chief: Krista Wilkinson
Editor: Preeti Sivasankar
Received March 9, 2017
Revision received June 3, 2017
Accepted June 12, 2017
https://doi.org/10.1044/2017_AJSLP-17-0031

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Empirical Studies

Real-time, self-reported vocal status estimates have been collected multiple times throughout a day by various repeated-measure study designs or ambulatory voice monitors. These types of studies have shown changes in self-ratings during and after heavy voice use in participants with normal voices who were college students (Vintturi, Alku, Sala, Sihvo, & Vilkman, 2003), singers or actors (Carroll et al., 2006; Kitch & Oates, 1994; Kitch, Oates, & Greenwood, 1996; Welham & Maclagan, 2003a), teachers (Gotaas & Starr, 1993; Halpern, Spielman, Hunter, & Titze, 2009; Hunter & Titze, 2009; Laukkanen, Ilomäki, Leppänen, & Vilkman, 2008; Laukkanen & Kankare, 2006; Popolo, Titze, & Hunter, 2011), and other groups of participants in high–voice-use occupations (Dallaston & Rumbach, 2016; Lehto, Laaksonen, Vilkman, & Alku, 2008). Furthermore, differences in voice use have been noted when comparing occupational versus nonoccupational time (Hunter & Titze, 2010), implying that overall vocal status may be worse during the workweek and may improve during the weekend. Because ambulatory voice monitors unobtrusively record phonation in these real-life contexts, subjective self-reported vocal status changes can be linked to or associated with various objective voice measures (Carroll et al., 2006; Halpern et al., 2009; Mehta et al., 2015; Verdyuckt, Rungassamy, Remacle, & Dubuisson, 2011). The resulting combination of objective and subjective data in this approach may provide insight into the ways in which phonatory behaviors lead to and/or interact with voice changes over the course of days or longer time scales. Furthermore, ambulatory voice monitors can cue the wearer when self-report questions need to be answered, which increases compliance compared with simple paper-and-pencil methods that are prone to being forgotten (Mehta, Cheyne, Wehner, Heaton, & Hillman, 2016).

Because all the previously cited ambulatory studies only tracked subjects with normal voices, results may not accurately represent pathological vocal status changes but instead provide insights into solely the normative (i.e., non-pathological) aspects of vocal deterioration and recovery. In addition, study designs using vocal loading tasks artificially elicit vocal status changes in participants with normal voices (Fujiki, Chapleau, Sundarrajan, McKenna, & Sivasankar, 2016; Remacle, Finck, Roche, & Morsomme, 2012; Whitling, Rydell, & Åhlander, 2015); therefore, results cannot be assumed to accurately represent pathological vocal conditions such as VH. For example, patients with VH—who are in chronic pathological vocal states—might not fluctuate in the same manner as those with normal voices. By way of analogy, simply studying how healthy subjects fatigue and recover with physical exertion would not be expected to provide insights into the disordered fatigue state associated with chronic fatigue characterized by minimal or no recovery with rest (Fukuda et al., 1994). In addition, even if vocal loading tasks included patients with VH (Mathur, Pande, Sahgal, & Reclaimation, 2016; Remacle, Morsomme, Berrué, & Finck, 2012), findings are difficult to clinically interpret because the nature of the task relies on circular logic; that is, the construct under investigation (vocal fatigue) is defined by whatever objective changes are noted because of an artificial task (Nanjundeswaran et al., 2015).

Theoretical Viewpoints

As a starting point, there are similarities in self-reported symptoms of vocal deterioration between individuals with normal voices in high–voice-use occupations (Kitch & Oates, 1994) and symptoms reported by patients (Ahlander, Rydell, & Löfqvist, 2012) and clinicians treating those patients (Nanjundeswaran et al., 2015). More specifically, three interrelated classes of symptoms related to vocal status have been commonly reported: (a) a decreased ability to produce softer, higher-pitched voicing; (b) pain or discomfort in the throat, jaw, and/or neck; and (c) vocal fatigue or increased phonatory effort (for reviews, see Nanjundeswaran et al., 2015; Welham & Maclagan, 2003b).

Vocal fatigue/effort has been the most thoroughly researched of the three vocal symptom classes but is difficult to objectively or subjectively quantify due to variable conceptual definitions, heterogeneous healthy and/or patient samples within and across studies, and frequent reports of high intersubject variability (Nanjundeswaran et al., 2015; Solomon, 2008; Welham & Maclagan, 2003b). The multiple theoretical definitions/models of vocal fatigue have been derived from various perspectives including basic science (McCabe & Titze, 2002), applied clinical science (Nanjundeswaran et al., 2015; Solomon, 2008), and patient interview (Kitch & Oates, 1994). In a recent study that included a focus group of expert voice clinicians, consensus was that vocal effort and fatigue were synonymous (Nanjundeswaran et al., 2015); therefore, vocal fatigue will be the term used throughout the rest of the article and is defined as the perception of increased vocal effort and/or tiredness by the voice user (Solomon, 2008).

Discomfort is frequently cited in studies of self-reported symptoms from patients with voice disorders (Lopes, Cabral, & de Almeida, 2015; Mathieson, 1993; Woznicka, Niebudek-Bogusz, Kwiecien, Wiktorowicz, & Sliwinska-Kowalsk, 2012) and is defined as any perceptual correlate of nociceptive sensations in the muscles or tissues related to speech and/or voice. Difficulty producing soft, high-pitched phonation (D-SHP) has been used in the clinical literature as a method to improve identification of vocal fold tissue damage (Bastian, Keidar, & Verdolini-Marston, 1990). The technique has also been used to assist teachers in rating their vocal status throughout the workweek via the "inability to produce soft voice" method contained in the National Center for Voice and Speech's Self-Administered Vocal Rating (Hunter & Titze, 2008). D-SHP and discomfort have often been subsumed under vocal fatigue, but physical tissue characteristics unrelated to vocal fatigue have direct effects on both self-ratings. For example, even without the impact of vocal fatigue, large vocal fold nodules could contribute to a patient's D-SHP judgment. Furthermore, in another example, acute laryngitis from an infectious process (not vocal misuse/abuse) could directly affect both the patient's levels of discomfort (i.e., increased soreness and pain) and his or her D-SHP judgment.

To the authors' knowledge, there remains a need to investigate vocal status changes throughout daily life in a group of patients diagnosed with VH and compare ecologically valid, ambulatory, real-time self-ratings of vocal status between patients with VH and subjects with normal voices. Thus, the purpose of this study is to quantify these basic comparisons using ambulatory estimates of D-SHP, vocal discomfort, and vocal fatigue. This data set will first test the three questions' psychometric properties of reliability and sensitivity to change. For reliability, it is hypothesized that the three questions of D-SHP, discomfort, and fatigue will all strongly relate to the global concept of "vocal status" (established via internal consistency) and that each will uniquely contribute to this global construct (correlation coefficients below the threshold of multicollinearity). Then, to assess validity, two hypotheses will be tested: (a) patients with VH will self-report higher levels of D-SHP, vocal discomfort, and vocal fatigue than the normative sample, and (b) patients with VH will self-report lower levels of the three vocal symptom classes after voice therapy or laryngeal surgery relative to pretreatment levels.

Method

Participants

All participants were enrolled in a larger study investigating the clinical utility of smartphone-based ambulatory voice monitoring (Mehta et al., 2015). The study sample consisted of 84 patients with VH (vocal fold nodules: two men and 44 women, polyps: six men and nine women, muscle tension dysphonia: six men and 17 women) and 74 subjects with normal voices (13 men and 61 women). The average age (mean ± standard deviation [*SD*]) of the patient group was 32.8 ± 13.7 years, and that of the normative group was 31.1 ± 13.4 years. Diagnoses were based on a comprehensive team evaluation (laryngologist and speech-language pathologist [SLP]) at the Massachusetts General Hospital Center for Laryngeal Surgery and Voice Rehabilitation (MGH Voice Center) that included (a) the collection of a complete case history, (b) endoscopic imaging of the larynx, (c) application of the Voice-Related Quality of Life (V-RQOL) questionnaire (Hogikyan & Sethuraman, 1999), (d) an auditory–perceptual evaluation using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster, Gerratt, Verdolini Abbott, Barkmeier-Kraemer, & Hillman, 2009), and (e) aerodynamic and acoustic assessments of vocal function. There was no additional assessment of self-reported vocal fatigue outside the typical questions asked during the collection of the patient's case history and in the V-RQOL. Once patients were diagnosed with vocal fold nodules, polyps, or muscle tension dysphonia, study staff at the MGH Voice Center approached them using recruitment and consenting procedures approved by the MGH's Institutional Review Board. All consented patients were asked to recruit an age- (±5 years), gender-, and occupation-matched control with no history of voice disorders, which composed the normative group. If the patient was unable to find a control with a normal voice, the study staff sought a control through email and flyer advertisements approved by the MGH Institutional Review Board. The vocal status of all participants in the normative group was verified via an interview and a laryngeal stroboscopic examination. During the interview, the candidates for the normative group were specifically asked if they had any voice difficulties that affected their daily life, and an SLP evaluated the auditory–perceptual quality of their voices. If the candidate indicated voice difficulties or showed a non-normal voice quality, he or she was excluded from study enrollment and did not undergo a laryngeal stroboscopic examination.

Table 1 reports subscale scores for the self-reported V-RQOL and clinician-judged CAPE-V ratings for the participants in the patient group. V-RQOL scores are normalized ordinal ratings that lie between 0 and 100 (inclusive), with higher scores indicating a higher quality of life. CAPE-V scores are visual analog scale ratings that range from 0 to 100, with 0 indicating normality and 100 indicating extremely severe abnormality of a particular voice quality characteristic. Scores on both scales indicated that most participants exhibited mild-to-moderate dysphonia and/or V-RQOL impairments. The CAPE-V measurement for each patient came from one rater—the treating SLP's single rating during a routine clinical evaluation.

Data Collection

Subjects were monitored each day over the course of 1 week (7 days) using a smartphone-based ambulatory voice monitor called the Voice Health Monitor (VHM; Mehta, Zañartu, Feng, Cheyne, & Hillman, 2012). Although the VHM uses an accelerometer to record phonation, the scope of the current study does not include using measures

Table 1. Patients' self-reported quality-of-life impact due to their voice disorder using the Voice-Related Quality of Life (V-RQOL) subscales and the perceived qualities of their voice as judged by a speech-language pathologist using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) form.

Evaluation scale	Patient group
V-RQOL	
Social–emotional	70.0 (22.1)
Physical functioning	68.9 (21.8)
Total score	69.3 (19.3)
CAPE-V	
Overall severity	28.9 (17.2)
Roughness	18.6 (15.8)
Breathiness	11.3 (12.6)
Strain	19.8 (14.7)
Pitch	8.0 (12.2)
Loudness	6.1 (12.5)

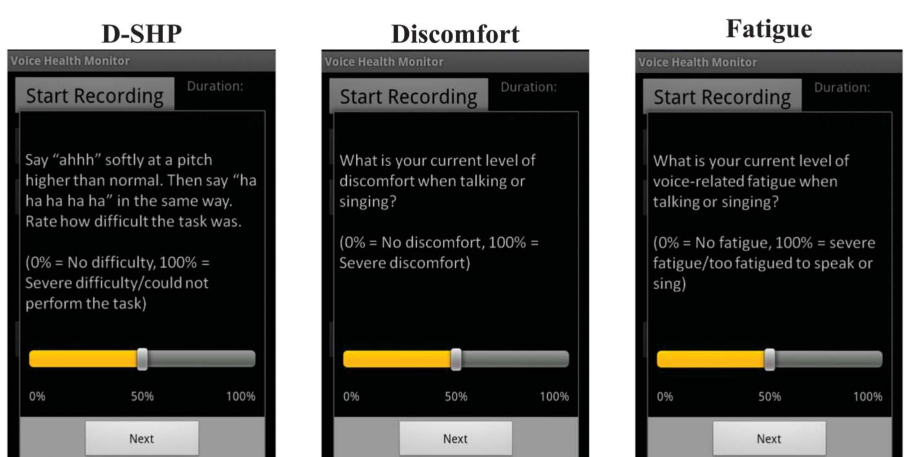
Note. Mean (*SD*) is reported for the patient group (*N* = 84).

extracted from the recorded accelerometer signal. During their consenting visit, all participants were thoroughly instructed on how to use the phone to respond to the three vocal status questions and given ample opportunity to have any remaining questions about the procedures answered by study staff. Figure 1 displays the three vocal status questions (D-SHP, vocal discomfort, and vocal fatigue) that the VHM prompts users to answer at the beginning and end of their day as well as at 5-hr intervals during the day. Users were offered the option to delay ("snooze") answering the questions by 5 min if prompts appeared at busy times. Questions were answered using a visual analog scale where the default placement was the middle (i.e., 50) of a 0–100 scale quantized at 0.25 increments.

To answer Question 1 (D-SHP), users were asked to hold an /a/ vowel for several seconds at a higher-than-normal pitch and softer-than-normal vocal intensity. Subsequently, they were asked to produce five staccato /ha/s at the same high pitch and soft intensity. The D-SHP gestures have been modified in two key ways compared with previous attempts at assessing soft, high-pitched voicing (Bastian et al., 1990; Halpern et al., 2009; Hunter & Titze, 2009). First, the gesture uses an /a/ vowel instead of an /i/ vowel (a carryover from laryngeal videostroboscopic examinations) because the higher first formant frequency (F_1) for /a/ minimizes vowel-related interactions with the high-pitched phonatory gesture (i.e., instability when f_o crosses F_1; Titze, 2008; Zañartu, Mehta, Ho, Wodicka, & Hillman, 2011). Second, the gesture has been reduced to one sustained and five staccato phonations to minimize testing burden (vs. the "inability to produce soft voice" method that calls for a much more extensive vocalization battery; Hunter & Titze, 2008). Question 2 (discomfort) asks the participants to rate any voice-related pain, soreness, or discomfort. Participants were specifically told that "discomfort" includes any pain or soreness. Question 3 (fatigue) asks the participants to rate the presence of vocal fatigue. Participants were specifically told that "fatigue" referenced any sensation of tiredness or increased effort required for voicing.

Two subsets of patients with VH completed a second week of monitoring after they successfully finished voice therapy or after finishing 2–3 weeks of voice rest post–laryngeal surgery to remove their phonotraumatic lesions. In the voice therapy group, participants were only monitored for a second week if their treatment was determined to be successful based on a comprehensive evaluation including acoustic, aerodynamic, and perceptual measures as well as subjective judgments of significant improvement from both the treating SLP and the patient. Note that the self-ratings of vocal status before and after therapy or surgery are being used to establish item validity, not to establish the three vocal status questions as outcome measures. Before the questions can be used to investigate clinical effectiveness or efficacy (e.g., outcomes measures), the items must statistically indicate simple hypothesized changes (e.g., levels of pathological vocal status decrease after routine clinical intervention). The first week of monitoring in all patients was before any type of clinical intervention. Because the focus of this study was investigating the statistical properties of self-rated estimates of vocal status, the procedures used for laryngeal surgeries and/or voice therapy sessions were not controlled for or documented in a standardized fashion. Measures were collected before and after treatment to better facilitate the chances of observing changes in the measures for statistical and validation purposes, not to determine the clinical efficacy/effectiveness of the treatment procedures. Each patient's voice therapy was individually tailored and typically included direct intervention treatment ingredients related to auditory, somatosensory, vocal function, musculoskeletal, and/or respiratory targets as well as indirect intervention treatment ingredients related to pedagogical and counseling targets (Van Stan, Roy, Awan, Stemple, & Hillman, 2015). There were 29 patients with VH who successfully completed voice therapy (18 with vocal

Figure 1. Vocal status questions as displayed on the smartphone-based ambulatory voice monitor for difficulty producing soft, high-pitched phonation (D-SHP; left); discomfort (center); and fatigue (right).



fold nodules, four with polyps, and seven with muscle tension dysphonia) and 16 patients with VH who underwent laryngeal surgery to remove their phonotraumatic lesions (10 with vocal fold nodules and six with polyps). Four of the 16 patients from the surgical group attempted voice therapy before deciding to undergo laryngeal surgery; thus, their 2 weeks of comparison were the week before surgery and the week after completing voice rest. The mean (*SD*) number of therapy sessions across the therapy group was 7 (4) sessions over the course of 79 (63) days.

Statistical Analysis

Reliability

Because all three aspects of vocal status were theorized to vary considerably throughout the day, test–retest reliability was not appropriate for these questions. Therefore, a measure of internal consistency, Cronbach's alpha, was the chosen metric of reliability (Tavakol & Dennick, 2011). More specifically, an alpha measures how closely all questions in the psychometric test load onto the same construct, with the construct being "vocal status." A degree of covariance among the three questions was expected, as they are all related and not considered to be orthogonal aspects of vocal status. However, if α was found to be greater than .9 (especially in the patient sample), this would potentially indicate too much redundancy among questions (Streiner et al., 2014). "Too much redundancy" among question responses would indicate that the questions do not individually provide much information and would warrant investigating which questions could be combined or removed. Cronbach's alpha was computed separately for the normative and patient groups because combining both data sets would increase the heterogeneity of the sample and produce an inflated alpha. Each participant was represented by the mean of their pooled ratings per question. Pairwise relationships between vocal symptom ratings were quantified using the coefficient of determination (r^2) to investigate the degree of overlap among questions.

Sensitivity to Change

The reliability coefficient (r) used to calculate sensitivity-to-change metrics was Cronbach's alpha from the patient group. Therefore, in addition to traditional statistical significance, clinically detectable differences could be examined (Guyatt, Walter, & Norman, 1987). More specifically, the standard error of measurement (*SEM*) was calculated by multiplying the individual question's standard deviation (from the patient-based ratings) by the square root of $(1 - r)$. The *SEM* thus provided an indication of the upper limit of variability inherent to each of the three test items. The *SEM* represents the smallest change the measure can realistically detect, which is much smaller than a clinically meaningful or detectable change. The minimal detectable change (MDC) is used to calculate the smallest change that is potentially clinically meaningful or clinically detectable. More specifically, the MDC using a 95% confidence interval (MDC_{95}) was calculated by multiplying the *SEM* by 1.96 (the 95th percentile of a standard normal distribution) and the square root of 2. The resulting number provided an indication of how much change in a rating must occur to be 95% certain that the change was not due to chance, which makes it very likely to be a clinically meaningful change.

Validity

If the question responses were truly providing an estimate of the three vocal symptom classes (D-SHP, vocal discomfort, and vocal fatigue), they should differentiate between the normative and patient samples. Therefore, each participant was represented by his or her mean rating for each of the three vocal symptom classes, and independent-samples t tests were used to determine if the patient sample reported significantly different ratings from those of the healthy sample. To examine any differences between subtypes of VH, independent-samples t tests (with equal variance assumed) were completed between patients with phonotraumatic VH (nodules and/or polyps) and patients with nonphonotraumatic VH (muscle tension dysphonia; see terminology in Mehta et al., 2015).

It was also hypothesized that patients should experience improved vocal status after successful voice therapy or laryngeal surgery. Pretherapy versus posttherapy and presurgery versus postsurgery analyses were completed using paired-samples t tests to determine if the patients significantly improved (i.e., reported reduced D-SHP, discomfort, and fatigue) posttreatment. Independent-samples t tests were used to determine if the patients "normalized" after treatment by comparing the posttherapy and postsurgery ratings with the normative ratings.

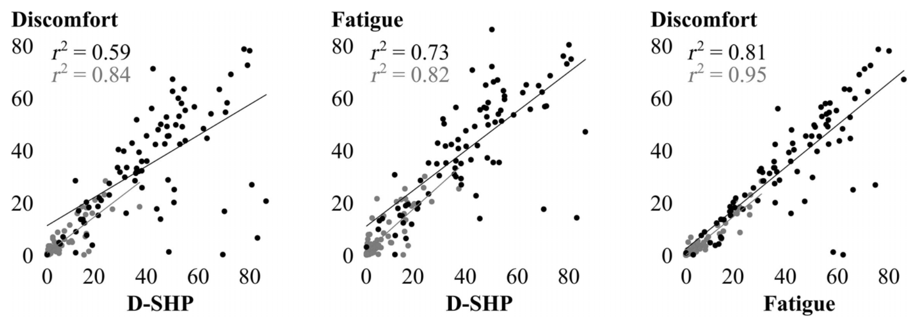
When significance was found for any statistic (Bonferroni-corrected p value threshold of .0167 due to three questions tested for each comparison), an associated effect size was determined using Cohen's d (i.e., the difference between the two groups' means divided by their pooled standard deviations). The effect size provided a standardized method to interpret the size of differences between the two groups, with $d < 0.2$ interpreted as small, $0.2 < d < 0.8$ interpreted as medium, and $d > 0.8$ interpreted as large (Cohen, 1988).

Results

Reliability

Cronbach's alpha for the group of patients with VH and the normative sample indicated high reliability (α = .88 and .95, respectively). Figure 2 summarizes the pairwise relationships between individual questions and between the two groups of participants. The patient group exhibited statistically significant pairwise relationships ($r^2 = .59$ for D-SHP vs. discomfort, $r^2 = .73$ for D-SHP vs. fatigue, $r^2 = .81$ for fatigue vs. discomfort), which were lower than the coefficients of determination from the normative group ($r^2 = .84$ for D-SHP vs. discomfort, $r^2 = .82$ for D-SHP vs. fatigue, $r^2 = .95$ for fatigue vs. discomfort).

Figure 2. Pairwise scatter plots of self-ratings of vocal status questions (patients with vocal hyperfunction: black, normative sample: gray). Each dot represents the mean of an individual's self-ratings pooled across 7 monitored days. The coefficient of determination (r^2) and best-fit regression line for each pairwise relationship are also displayed. D-SHP = difficulty producing soft, high-pitched phonation.



Sensitivity to Change

Using Cronbach's alpha derived from the patient data, the *SEM* values for D-SHP, discomfort, and fatigue were calculated as 7.14, 6.99, and 7.16, respectively. Using these *SEM* values, the MDC_{95} values for D-SHP, discomfort, and fatigue were computed to be 19.79, 19.37, and 19.85, respectively. Therefore, a patient would need to change approximately 20 points for one to conclude with 95% certainty that the vocal status change is clinically detectable and meaningful.

Validity

Table 2 summarizes the mean (*SD*) values across participants for each of the three self-ratings. Overall, there was a statistically significant difference between mean self-ratings for the group of patients with VH and the normative sample. Independent-samples *t* tests (equal variance was not assumed) showed statistically significant differences ($p < .001$) between patient and normative self-ratings of D-SHP, discomfort, and fatigue: $t(107) = 14.53$, $t(107) = 12.77$, and $t(107) = 14.17$, respectively. More specifically, the patients reported higher levels of pathological vocal status than the normative sample with large effect sizes: $d = 1.92$, 1.62, and 1.88 for D-SHP, discomfort, and fatigue, respectively. The *t* tests comparing D-SHP, discomfort, and fatigue between the subgroups with phonotraumatic and nonphonotraumatic VH were not statistically significant, indicating no differences between VH subtypes.

Table 3 summarizes the mean and standard deviation values across patient treatment levels (therapy or surgery) for each of the three self-ratings. Overall, paired *t* tests showed statistically significant differences between mean self-ratings before and after therapy in all questions: D-SHP, $t(28) = 3.76$, $p < .001$; discomfort, $t(28) = 4.76$, $p < .001$; and fatigue, $t(28) = 3.91$, $p < .001$. Compared with their pretherapy ratings, patients reported lower levels of pathological vocal status after successful therapy with medium-to-large effect sizes ($d = 0.70$, 0.88, and 0.85 for D-SHP, discomfort, and fatigue, respectively). All group-based differences pretherapy/posttherapy were well above the *SEM* and only slightly below the MDC_{95}. Paired *t* tests also showed statistically significant differences between mean ratings before and after laryngeal surgery in all three questions: D-SHP, $t(15) = 4.54$, $p < .001$; discomfort, $t(15) = 3.26$, $p = .005$; and fatigue, $t(15) = 4.42$, $p < .001$. The patients reported lower levels of pathological vocal status after laryngeal surgery compared with before surgery with large effect sizes ($d = 1.14$, 0.81, and 1.11 for D-SHP, discomfort, and fatigue, respectively). All group-based differences presurgery/postsurgery were well above the MDC_{95}. Although both patient groups significantly improved after voice therapy or laryngeal surgery, independent-samples *t* tests (with no assumption of equal variance) indicated that these patient groups still reported significantly higher levels of pathological vocal status than reported by the normative group posttherapy [D-SHP, $t(28) = 3.17$, $p = .003$, $d = 0.81$; discomfort, $t(28) = 3.22$, $p = .003$, $d = 0.83$; and fatigue, $t(28) = 3.05$, $p = .004$, $d = 0.78$] and postsurgery [D-SHP,

Table 2. Group-based mean (*SD*) for self-reported difficulty producing soft, high-pitched phonation (D-SHP); discomfort; and fatigue during 7 days of ambulatory voice monitoring reported within the patient group ($N = 84$), nodules and polyps group ($n = 61$), muscle tension dysphonia group ($n = 23$), and normative group ($N = 74$).

Vocal status questions	Patients with vocal hyperfunction			Normative group
	All	Nodules and polyps	Muscle tension dysphonia	
D-SHP	40.99 (20.69)	43.18 (20.58)	35.20 (20.28)	5.79 (7.54)
Discomfort	34.87 (20.25)	36.89 (20.01)	29.51 (20.35)	4.99 (6.61)
Fatigue	41.16 (20.76)	42.52 (20.32)	37.55 (21.94)	6.42 (8.05)

Table 3. Group-based mean (*SD*) for self-reported difficulty producing soft, high-pitched phonation (D-SHP); discomfort; and fatigue during approximately 7 days of ambulatory voice monitoring reported within patient subgroups who underwent therapy (n = 29) or laryngeal surgery (n = 16).

Vocal status questions	Patients with vocal hyperfunction			Nodules and polyps		
	Before therapy	After therapy	Difference	Before surgery	After surgery	Difference
D-SHP	31.28 (17.69)	18.23 (20.24)	–13.05 (18.69)	59.31 (20.05)	26.58 (19.61)	–32.72 (28.82)
Discomfort	32.84 (20.74)	16.45 (18.41)	–16.39 (18.55)	45.19 (20.53)	23.29 (25.16)	–21.90 (26.90)
Fatigue	36.20 (21.66)	18.38 (20.16)	–17.82 (21.06)	55.16 (19.30)	26.18 (23.38)	–28.98 (26.21)

Note. Values in "Difference" columns are computed on within-subject changes in ratings.

$t(15) = 4.17$, $p < .001$, $d = 1.40$; discomfort, $t(15) = 2.89$, $p = .011$, $d = 1.00$; and fatigue, $t(15) = 3.34$, $p = .004$, $d = 1.13$].

Discussion

Reliability

One objective of this study was to establish psychometric properties of reliability for the ambulatory vocal status questions in a large group of patients with VH and subjects with normal voices. The questions maintained high reliability in both groups of participants, with the normative sample resulting in higher Cronbach's alpha and question-versus-question coefficients of determination. One potential reason for the higher values is likely due to the limited variation exhibited in the normative group. In Figure 2, the patient data have a noticeably larger variance compared with the normative data. A second potential reason could be that the questions are too redundant when answered by participants with normal voices; that is, when participants are not experiencing pathological vocal status, the three items/symptoms characterizing vocal status do not differentiate themselves. All coefficients of determination using the normal data are well above the standard r^2 cutoff of .80 for the statistical concept of multicollinearity, and Cronbach's α above .90 implies that the three questions may have too much redundancy (Streiner et al., 2014). However, in contrast to the data from participants with normal voices, the patient data showed partial independence among questions, with only one pairwise question relationship (discomfort vs. fatigue) passing the threshold of multicollinearity ($r^2 = .80$) with an r^2 of .81. Furthermore, the patient-based Cronbach's alpha remained below the threshold of excess redundancy ($\alpha = .90$), with an α of .88. Perhaps the three aspects of vocal status outlined in the literature only achieve a sense of partial independence in a pathological vocal condition, such as VH. In other words, perhaps, people with VH more commonly self-report changes in one or two questions independently (e.g., increases in fatigue and D-SHP yet no self-reported discomfort). In contrast, people with normal voices simply self-report overall changes in healthy vocal status (which do not attain severity levels that permit symptom divergence). Because the degree of redundancy among questions can be a potential predictor of the presence or absence of a voice disorder, it is not recommended to collapse the three questions into one global question of vocal status.

Sensitivity to Change

A second objective of this study was to use the reliability data from the patients with VH to obtain sensitivity-to-change estimates such as the *SEM* and MDC_{95}. Considering the overall standard deviation of the resulting data (approximately 20 points), an approximate *SEM* of 7 and an MDC_{95} of 20 appear to be reasonable for clinical and research use. For example, the questions can measure small standardized effect sizes as low as .35 (*SEM* divided by the standard deviation), which could be useful for quantifying changes in participants with normal voices who typically are not expected to exhibit large changes from pathological to healthy vocal status. Furthermore, when clinically meaningful changes are expected in patients with VH due to a surgery or voice therapy, a standardized effect size of 1.0 (MDC_{95} divided by the standard deviation) is not unreasonably large. Finally, the average intraindividual variability (measured via standard deviation) provides insight into how much vocal status variation occurred within an individual. More specifically, the mean individual patient standard deviation was approximately 15 points (1 standard deviation across all patients was ±7 points), and the mean individual healthy control standard deviation was 6 points (1 standard deviation across all controls was ±5 points), meaning that most daily/weekly changes in patient-reported vocal status exceeded the *SEM* for the three items but fell short of the threshold for being minimally detectable.

Validity

A third objective was to test the validity of the three questions through comparisons of mean ratings between patients with VH and a normative sample. The patients with VH reported significantly higher levels of D-SHP, discomfort, and fatigue compared with those from the normative sample with large effect sizes. Furthermore, there were no differences between subtypes of VH; that is, patients with phonotraumatic and nonphonotraumatic VH appeared to show equivalent levels of pathological vocal status. Although D-SHP, discomfort, and fatigue are all commonly reported symptoms in both diagnoses, it was surprising that D-SHP in particular was not higher in the group with phonotraumatic VH compared with the group with nonphonotraumatic VH. The soft, high-pitched gesture has commonly been used to help perceptually identify or rate true vocal fold tissue changes (Bastian et al., 1990; Hunter & Titze, 2008), which would suggest that the gesture could differentiate those with vocal fold tissue pathology (phonotraumatic VH) from those without vocal fold tissue pathology (nonphonotraumatic VH). However, the fundamental frequency and amplitude of the vocal gesture were not examined in this article, and perhaps, a difference between the VH subtypes may yet be found in future investigations that include those two objective measures; for example, perhaps the measures could be added as covariates to the statistical analyses.

To investigate the validity of the MDC_{95} (i.e., Will it adequately detect within-patient changes?), 77 patients with VH were included in an additional analysis where a patient's mean rating from the beginning of the day was compared with his or her mean rating at the end of the day (seven patients were excluded because they did not have consistent morning and evening data per day due to technical difficulties, snoozing, or decreased compliance). Then, using the MDC_{95} as a cutoff, it became possible to identify how many patients had worse vocal symptoms at the beginning versus at the end of the day or no difference. The prevalence of patients who experienced "worse symptoms" at particular time points was compared with that of a larger sample of patients with VH ($N = 2{,}592$) extracted from the MGH Voice Center clinical database. Every patient who is treated at the MGH Voice Center is asked if his or her vocal symptoms are worse at various time points during a day, and the resulting data are summarized in Table 4. The MDC_{95} appears to be of adequate sensitivity because it identified a similar percentage of patients reporting varying symptoms (25%) as the larger sample (23%).

Interestingly, 75%–80% of patients with VH reported minimal or no variation of their vocal symptoms throughout a day. Although this is in line with our own clinical data reported above, it appears at odds with most studies and theories investigating vocal fatigue, where the central concept of vocal loading assumes vocal recovery with voice rest and vocal deterioration with voice use (Carroll et al., 2006; Gotaas & Starr, 1993; Halpern et al., 2009; Hunter & Titze, 2009; Kitch & Oates, 1994; Kitch et al., 1996; Laukkanen et al., 2008; Laukkanen & Kankare, 2006; Popolo et al., 2011; Welham & Maclagan, 2003a). Maybe pathological vocal status and changes in status are not similar to voice changes induced or mimicked in studies of participants with normal voices—perhaps vocal symptoms reported by patients are much more chronic in nature. To begin testing these assumptions, future analyses could incorporate measures of voice use (e.g., vocal doses) and vocal function (objective voice measures) to investigate the relationship between objective measures and self-reported status, similar to some previous studies (Carroll et al., 2006; Hunter & Titze, 2009). The analyses in this study only investigated self-reported data and cannot answer questions regarding relationships between objective and subjective measures of vocal function.

Finally, validity testing also included comparisons before and after two treatment conditions (voice therapy and laryngeal surgery), where the questions were hypothesized to show significant decreases posttreatment in both cases. The patients with VH reported significantly lower levels of all three constructs posttherapy and postsurgery, and the effect sizes were large. Because all previous self-reported comparisons preintervention/postintervention have been one-time surveys completed at the voice clinic (for samples, see Hogikyan & Sethuraman, 1999; Jacobson et al., 1997; Nanjundeswaran et al., 2015), this finding in itself (large changes in self-reported vocal status during daily life) is novel for the field of voice therapy. When applying the MDC_{95}, surgery-related changes in all three ratings far exceeded the 20-point cutoff, whereas the voice therapy–related changes were slightly below the cutoff. One reason for this could be the lower level of severity in the patients before therapy (a group that contained patients with phonotraumatic and nonphonotraumatic VH) compared with the patients before surgery (a group of only patients with phonotraumatic VH); that is, the patients in the pretherapy subgroup had less room to improve compared with the patients in the presurgery subgroup. The most striking difference between the surgical and voice therapy changes related to the D-SHP question. More specifically, the largest

Table 4. Number of patients reporting vocal symptom changes (or no changes) throughout their daily life.

When do vocal symptoms worsen?	Vocal status change ≥ 20 points (MDC_{95})	MGH Voice Center clinical database		
		All patients with VH	Nodules and polyps	Muscle tension dysphonia
Beginning of the day	4	287	179	108
End of the day	15	245	142	103
Other	—	237	110	127
No variation reported	58 (75%)	2,592 (77%)	1,068 (71%)	1,524 (82%)

Note. Column 2 shows patients from the current study who reported ≥ 20 points (minimal detectable change using 95% confidence intervals [MDC_{95}]) in difficulty producing soft, high-pitched phonation; discomfort; and/or fatigue when comparing the mean "beginning of the day" and "end of the day" ratings. Columns 3–5 show data from the Massachusetts General Hospital (MGH) Voice Center clinical database where patients with vocal hyperfunction (VH) self-reported if and when their vocal symptoms worsened throughout the course of a day.

treatment-related changes for the surgical group occurred with the D-SHP measure, whereas the therapy group exhibited the smallest treatment-related changes with this question. This result is most likely due to the removal of the true vocal fold lesions, which maximized glottal closure, normalized the physics related to phonation, and therefore made the D-SHP task less difficult as would be expected (Zeitels, Hillman, Desloge, Mauri, & Doyle, 2002). Finally, compared with the normative sample, patients were still reporting higher levels of all three symptoms postsurgery and posttherapy. The postsurgery finding was expected because vocal behavior is believed to play a major role in the etiology of nodules and polyps and surgery does not directly address these potentially habitual behaviors. However, the posttherapy finding suggests that patients with VH achieve a significant improvement in vocal function that is adequate for their vocal needs but may not completely return to "normal."

Clinical Implications

The immediate clinical significance of this study is the establishment of three theoretically and empirically validated questions for periodically assessing the vocal status of patients with hyperfunctional voice disorders during activities of daily living. In the current study, an app on a smartphone was used to facilitate data collection; however, clinicians could ask patients to respond to the three questions throughout daily life simply using paper and pencil. The sensitivity-to-change metrics provide useful thresholds for interpreting changes in responses over time; that is, a change of 20 points or more is considered clinically meaningful.

The three vocal status questions are viewed as complementary to other statistically validated patient self-report instruments that are administered in the clinic and designed to assess the overall functional, emotional, and/or psychosocial impact of voice disorders in daily life (cf. Hogikyan & Sethuraman, 1999; Jacobson et al., 1997; Nanjundeswaran et al., 2015). Repeated ambulatory self-assessments of vocal status during patients' normal daily activities may provide a means to more accurately pinpoint periods when behaviors or situations are associated with deterioration in specific aspects of vocal status (e.g., particular communication or environmental situations). Clinicians could use this information to help patients develop more focused strategies for mitigating the negative impact of such situations on vocal status and as more sensitive metrics for determining whether the chosen strategies are effective. It is also possible that ambulatory self-assessments may be inherently more valid, reliable, and sensitive than instruments used in the clinic because the immediacy of the responses relies less on global judgments and long-term memory.

A primary motivation for validating the subjective ambulatory self-assessment questions is so they can be used as a basis for further developing ambulatory voice monitoring as a clinical tool (Mehta et al., 2012). Future work will seek to identify objective ambulatory measures of vocal function that are maximally sensitive to patient self-reported changes in vocal status. This is part of an effort to develop more quantitatively based metrics for defining healthy versus damaging/abnormal vocal function and thus improve the prevention, diagnosis, and treatment of hyperfunctional voice disorders (Hillman et al., 1989; Hillman, Heaton, Masaki, Zeitels, & Cheyne, 2006). Of particular interest in this regard is the identification of patient-specific objective measures appropriate for ambulatory voice biofeedback—which would theoretically improve carryover of therapeutic behaviors into the patients' daily lives (Schalling, Gustafsson, Ternström, Bulukin Wilén, & Södersten, 2013; Van Stan, Mehta, & Hillman, 2015; Van Stan, Mehta, Petit, et al., 2017; Van Stan, Mehta, Sternad, Petit, & Hillman, 2017).

Conclusions

Ambulatory self-estimates of D-SHP, vocal discomfort, and vocal fatigue indicated adequate reliability to detect vocal status changes in patients with VH as well as in participants with normal vocal function. Furthermore, the data obtained from these self-estimates exhibited the hypothesized differences between patients with VH and participants with normal voices (higher ratings by patients) as well as reduced patient self-ratings postsurgery and posttherapy. Because self-rating data can be collected during the course of ambulatory voice monitoring, future studies could investigate correlations between subjective ratings and objective measures of voice use (e.g., Will ratings increase as percentage of phonation increases?) or evaluate differences in objective voice measures between time points where the patient reports a worsening, no change, or an improvement of vocal status (given the *SEM* and MDC_{95} of 7 and 20 points, respectively).

Acknowledgments

This work was supported by the Voice Health Institute (Boston, MA), the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders under Grant R33 DC011588 (awarded to Dr. Hillman), and the National Counsel of Technological and Scientific Development from Brazil (CNPq–Conselho Nacional de Desenvolvimento Científico e Tecnológico; scholarship awarded to Dr. Masson). The article's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References

Ahlander, V. L., Rydell, R., & Löfqvist, A. (2012). How do teachers with self-reported voice problems differ from their peers with self-reported voice health? *Journal of Voice, 26*(4), e149–e161.

Bastian, R. W., Keidar, A., & Verdolini-Marston, K. (1990). Simple vocal tasks for detecting vocal fold swelling. *Journal of Voice, 4*(2), 172–183.

Carroll, T., Nix, J., Hunter, E., Emerich, K., Titze, I., & Abaza, M. (2006). Objective measurement of vocal fatigue in classical singers: A vocal dosimetry pilot study. *Otolaryngology-Head and Neck Surgery, 135*(4), 595–602.

Chung, J. H., Tae, K., Lee, Y. S., Jeong, J. H., Cho, S. H., Kim, K. R., . . . Han, D. S. (2009). The significance of laryngopharyngeal reflux in benign vocal mucosal lesions. *Otolaryngology-Head and Neck Surgery, 141*(3), 369–373.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Dallaston, K., & Rumbach, A. F. (2016). Vocal performance of group fitness instructors before and after instruction: Changes in acoustic measures and self-ratings. *Journal of Voice, 30*(1), 127.e121–127.e128.

Dietrich, M., Andreatta, R. D., Jiang, Y., Joshi, A., & Stemple, J. C. (2012). Preliminary findings on the relation between the personality trait of stress reaction and the central neural control of human vocalization. *International Journal of Speech-Language Pathology, 14*(4), 377–389.

Dietrich, M., Verdolini Abbott, K., Gartner-Schmidt, J., & Rosen, C. A. (2008). The frequency of perceived stress, anxiety, and depression in patients with common pathologies affecting voice. *Journal of Voice, 22*(4), 472–488.

Fujiki, R. B., Chapleau, A., Sundarrajan, A., McKenna, V., & Sivasankar, M. P. (2016). The interaction of surface hydration and vocal loading on voice measures. *Journal of Voice, 31*(2), 211–217.

Fukuda, K., Straus, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G., & Komaroff, A. (1994). The chronic fatigue syndrome: A comprehensive approach to its definition and study. *Annals of Internal Medicine, 121*(12), 953–959.

Gotaas, C., & Starr, C. D. (1993). Vocal fatigue among teachers. *Folia Phoniatrica et Logopaedica, 45*(3), 120–129.

Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases, 40*(2), 171–178.

Halpern, A. E., Spielman, J. L., Hunter, E. J., & Titze, I. R. (2009). The inability to produce soft voice (IPSV): A tool to detect vocal change in school-teachers. *Logopedics, Phoniatrics, Vocology, 34*(3), 117–127.

Hillman, R. E., Heaton, J. T., Masaki, A., Zeitels, S. M., & Cheyne, H. A. (2006). Ambulatory monitoring of disordered voices. *Annals of Otology, Rhinology and Laryngology, 115*(11), 795–801.

Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech and Hearing Research, 32*(2), 373–392.

Hogikyan, N. D., & Sethuraman, G. (1999). Validation of an instrument to measure voice-related quality of life (V-RQOL). *Journal of Voice, 13*(4), 557–569.

Hunter, E. J., & Titze, I. (2008). *General statistics of the NCVS self-administered vocal rating (SAVRa)*. Retrieved from http://www.ncvs.org/e-learning/tech/tech-memo-11.pdf

Hunter, E. J., & Titze, I. R. (2009). Quantifying vocal fatigue recovery: Dynamic vocal recovery trajectories after a vocal loading exercise. *Annals of Otology, Rhinology and Laryngology, 118*(6), 449–460.

Hunter, E. J., & Titze, I. R. (2010). Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings. *Journal of Speech, Language, and Hearing Research, 53*(4), 862–875.

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech-Language Pathology, 6*(3), 66–70.

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory–perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*(2), 124–132.

Kitch, J. A., & Oates, J. (1994). The perceptual features of vocal fatigue as self-reported by a group of actors and singers. *Journal of Voice, 8*(3), 207–214.

Kitch, J. A., Oates, J., & Greenwood, K. (1996). Performance effects on the voices of 10 choral tenors: Acoustic and perceptual findings. *Journal of Voice, 10*(3), 217–227.

Laukkanen, A.-M., Ilomäki, I., Leppänen, K., & Vilkman, E. (2008). Acoustic measures and self-reports of vocal fatigue by female teachers. *Journal of Voice, 22*(3), 283–289.

Laukkanen, A. M., & Kankare, E. (2006). Vocal loading-related changes in male teachers' voices investigated before and after a working day. *Folia Phoniatrica et Logopaedica, 58*(4), 229–239.

Lehto, L., Laaksonen, L., Vilkman, E., & Alku, P. (2008). Changes in objective acoustic measurements and subjective voice complaints in call center customer-service advisors during one working day. *Journal of Voice, 22*(2), 164–177.

Lopes, L. W., Cabral, G. F., & de Almeida, A. A. F. (2015). Vocal tract discomfort symptoms in patients with different voice disorders. *Journal of Voice, 29*(3), 317–323.

Mathieson, L. (1993). Vocal tract discomfort in hyperfunctional dysphonia. *Journal of Voice, 2,* 40–48.

Mathur, R. P., Pande, K., Sahgal, L., & Reclaimation, B. (2016). Impact of vocal loading test in primary school teachers with and without dysphonia. *International Journal of Health Sciences and Research, 6*(2), 328–334.

McCabe, D. J., & Titze, I. R. (2002). Chant therapy for treating vocal fatigue among public school teachers: A preliminary study. *American Journal of Speech-Language Pathology, 11*(4), 356–369.

Mehta, D. D., Cheyne, H. A., Wehner, A., Heaton, J. T., & Hillman, R. E. (2016). Accuracy of self-reported estimates of daily voice use in adults with normal and disordered voices. *American Journal of Speech-Language Pathology, 25*(4), 634–641.

Mehta, D. D., Van Stan, J. H., Zañartu, M., Ghassemi, M., Guttag, J. V., Espinoza, V. M., . . . Hillman, R. E. (2015). Using ambulatory voice monitoring to investigate common voice disorders: Research update. *Frontiers in Bioengineering and Biotechnology, 3,* 155.

Mehta, D. D., Zañartu, M., Feng, S. W., Cheyne, H. A., II, & Hillman, R. E. (2012). Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. *IEEE Transactions on Biomedical Engineering, 59*(11), 3090–3096.

Nanjundeswaran, C., Jacobson, B. H., Gartner-Schmidt, J., & Abbott, K. V. (2015). Vocal Fatigue Index (VFI): Development and validation. *Journal of Voice, 29*(4), 433–440.

Popolo, P. S., Titze, I. R., & Hunter, E. J. (2011). Towards a self-rating tool of the inability to produce soft voice based on nonlinear events: A preliminary study. *Acta Acustica, 97*(3), 373–381.

Remacle, A., Finck, C., Roche, A., & Morsomme, D. (2012). Vocal impact of a prolonged reading task at two intensity levels: Objective measurements and subjective self-ratings. *Journal of Voice, 26*(4), e177–e186.

Remacle, A., Morsomme, D., Berrué, E., & Finck, C. (2012). Vocal impact of a prolonged reading task in dysphonic versus normophonic female teachers. *Journal of Voice, 26*(6), 820.e81–820.e813.

Schalling, E., Gustafsson, J., Ternström, S., Bulukin Wilén, F., & Södersten, M. (2013). Effects of tactile biofeedback by a portable voice accumulator on voice sound level in speakers with Parkinson's disease. *Journal of Voice, 27*(6), 729–737.

Solomon, N. P. (2008). Vocal fatigue and its relation to vocal hyperfunction. *International Journal of Speech-Language Pathology, 10*(4), 254–266.

Streiner, D. L., Norman, G. R., & Cairney, J. (2014). *Health measurement scales: A practical guide to their development and use*. New York, NY: Oxford University Press.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2,* 53–55.

Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: Theory. *The Journal of the Acoustical Society of America, 123*(5), 2733–2749.

Van Stan, J. H., Mehta, D. D., & Hillman, R. E. (2015). The effect of voice ambulatory biofeedback on the daily performance and retention of a modified vocal motor behavior in participants with normal voices. *Journal of Speech, Language, and Hearing Research, 58*(3), 713–721.

Van Stan, J. H., Mehta, D. D., Petit, R. J., Sternad, D., Muise, J., Burns, J. A., & Hillman, R. E. (2017). Integration of motor learning principles into real-time ambulatory voice biofeedback and example implementation via a clinical case study with vocal fold nodules. *American Journal of Speech-Language Pathology, 26*(1), 1–10.

Van Stan, J. H., Mehta, D. D., Sternad, D., Petit, R., & Hillman, R. E. (2017). Ambulatory voice biofeedback: Relative frequency and summary feedback effects on performance and retention of reduced vocal intensity in the daily lives of participants with normal voices. *Journal of Speech, Language, and Hearing Research, 60*(4), 853–864.

Van Stan, J. H., Roy, N., Awan, S., Stemple, J., & Hillman, R. E. (2015). A taxonomy of voice therapy. *American Journal of Speech-Language Pathology, 24*(2), 101–125.

Verdyuckt, I., Rungassamy, C., Remacle, M., & Dubuisson, T. (2011). Real-time embedded tracking of patient reported vocal discomfort in professional settings. *Proceedings of the 7th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications* (pp. 157–160). Firenze, Italy: Firenze University Press.

Vintturi, J., Alku, P., Sala, E., Sihvo, M., & Vilkman, E. (2003). Loading-related subjective symptoms during a vocal loading test with special reference to gender and some ergonomic factors. *Folia Phoniatrica et Logopaedica, 55*(2), 55–69.

Welham, N. V., & Maclagan, M. A. (2003a). Vocal fatigue in young trained singers across a solo performance: A preliminary study. *Logopedics, Phoniatrics, Vocology, 29*(1), 3–12.

Welham, N. V., & Maclagan, M. A. (2003b). Vocal fatigue: Current knowledge and future directions. *Journal of Voice, 17*(1), 21–30.

Whitling, S., Rydell, R., & Åhlander, V. L. (2015). Design of a clinical vocal loading test with long-time measurement of voice. *Journal of Voice, 29*(2), 261.e13–261.e27.

Woznicka, E., Niebudek-Bogusz, E., Kwiecien, J., Wiktorowicz, J., & Sliwinska-Kowalsk, M. (2012). Applicability of the vocal tract discomfort (VTD) scale in evaluating the effects of voice therapy of occupational voice disorders. *Medycyna Pracy, 63*(2), 141–152.

Zañartu, M., Mehta, D. D., Ho, J. C., Wodicka, G. R., & Hillman, R. E. (2011). Observation and analysis of *in vivo* vocal fold tissue instabilities produced by nonlinear source-filter coupling: A case study. *The Journal of the Acoustical Society of America, 129*(1), 326–339.

Zeitels, S. M., Hillman, R. E., Desloge, R., Mauri, M., & Doyle, P. (2002). Phonomicrosurgery in singers and performing artists: Treatment outcomes, management theories, and future-directions. *Annals of Otology, Rhinology and Laryngology, 111,* 21–40.