Patient-Perceived Vocal Effort Immediately Following Voice Tasks in Adductor Laryngeal Dystonia (ADLD)

[*]Cara L. Sauder, [†,‡]Katherine L. Marks, [§]Tanya K. Meyer, [§]John Paul Giliberto, [§]Madeline Knutson, [§]Emily Wilson, [†,¶,‖]Cara E. Stepp, and [*,†]Tanya L. Eadie, *§*Seattle, Washington, †¶‖Boston, Massachusetts, and ‡Atlanta, Georgia*

Summary: Purpose. This study examined the relationship between patient-perceived vocal effort (VE) using a 100-mm visual analog scale (VE-VAS) and the OMNI Vocal Effort Scale (OMNI-VES) when measures were obtained after a vocal activity. A second purpose was to evaluate how VE related to other voice assessment measures.

Method. Fifty-three speakers with adductor laryngeal dystonia (ADLD) provided speech recordings. Directly after this vocal activity, speakers rated VE using the VE-VAS and the OMNI-VES. Speakers provided ratings of their own voice quality severity using a 100-mm VAS (ADLD-OS) and completed the Voice-Related Quality of Life (V-RQOL) scale. Ten experienced speech-language pathologists rated a subset of available speech samples (n = 39) for overall voice severity using a 100-mm VAS (SLP-OS).

Results. There was a strong, significant correlation ($r = 0.78$, $P < 0.001$) between the VE-VAS and the OMNI-VES. Both VE measures were strongly and significantly correlated with speakers' ratings of their voice: VE-VAS vs ADLD-OS, $r = 0.75$, $P < 0.001$; OMNI-VES vs ADLD-OS, $r = 0.85$, $P < 0.001$. In contrast, mostly weak correlations were found between perceived VE and V-RQOL total and physical domains, respectively (VE-VAS vs V-RQOL: $r = -0.21$ to -0.19, $P > 0.05$; OMNI-VES vs V-RQOL: $r = -0.37$ to -0.44, $P < 0.01$). Finally, VE measures were moderately and significantly related to SLPs' auditory-perceptual measures of voice severity: VE-VAS vs SLP-OS, $r = 0.50$, $P < 0.001$; OMNI-VES vs SLP-OS, $r = 0.57$, $P < 0.001$.

Conclusions. ADLD speakers' perceptions of VE are strongly related when measures are obtained directly after a vocal activity, regardless of the VE scale. VE is strongly related to speaker-rated voice quality severity, but weakly related to V-RQOL. Measures of VE obtained directly after a vocal activity are moderately related to clinicians' perceptions of overall voice quality severity.

Key Words: Laryngeal dystonia—Voice assessment—Vocal effort.

INTRODUCTION

Adductor laryngeal dystonia (ADLD) is a rare, neurologic movement disorder that is characterized by involuntary and intermittently forceful adduction of the intrinsic adductor laryngeal muscles during purposeful speech.[1,2] Speakers with ADLD are often perceived by listeners as having poor voice quality, frequently characterized by vocal strain as well as intermittent pitch and voice breaks.[3–8] Speakers with ADLD often report difficulty communicating with others, due in part to their altered voice quality and the undependability of their voices. However, perhaps the most prominent physical symptom reported by speakers with ADLD relates to their perception of increased vocal effort (VE),[9] defined as the perceived work or exertion an individual feels during phonation during speech tasks.[10–12] Among other factors, both poor voice quality and increased VE contribute to the negative physical, functional, and social-emotional consequences commonly experienced by speakers with ADLD, as measured using patient-reported outcome measures.[8,13]

The current "gold standard" and most common treatment for ADLD is botulinum toxin (BTX).[14] For many speakers with ADLD, BTX temporarily improves voice quality and reduces VE, resulting in improved Voice-Related Quality of Life (V-RQOL)[15] scores and increased communicative participation.[6,13,14,16,17] Clinicians also consider multiple factors, such as improved voice quality and reduced VE, as well as patient preferences, such as tolerance of side effects (eg, common breathy period), when they make decisions about the optimal dosage and frequency of BTX injections in patients with ADLD.[14,18,19] However, BTX is not effective for all speakers, and although BTX can provide symptomatic relief, speakers with ADLD must manage living with ADLD for the rest of their lives.[9,14,16,18,19] As a result, it is important that clinicians are able to validly assess and track symptoms that are meaningful to speakers with ADLD, such as perceived VE, and understand how this symptom might relate to other measures that are commonly obtained as part of a multidimensional voice assessment protocol.[5,7,8,10,11,20,21]

Accepted for publication November 12, 2024.

Portions of this paper were presented at the 53rd Annual Symposium: Care of the Professional Voice, June 2024, Philadelphia, PA.

From the *Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington; †Department of Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts; ‡Department of Otolaryngology-Head and Neck Surgery, Emory University, Atlanta, Georgia; §Department of Otolaryngology-Head and Neck Surgery, University of Washington, Seattle, Washington; ¶Department of Otolaryngology-Head and Neck Surgery, Boston University, Boston, Massachusetts; and the ‖Department of Biomedical Engineering, Boston University, Boston, Massachusetts.

Address correspondence and reprint requests to Cara L. Sauder, Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd Street, Seattle, WA 98105. E-mail: csauder2@uw.edu

Journal of Voice, Vol xx, No xx, pp. xxx–xxx

0892-1997

https://doi.org/10.1016/j.jvoice.2024.11.022

Yet, a recent systematic review[13] of treatment studies in laryngeal dystonia reported that among the 34 studies (n = 34/125, 27%) that used patient self-evaluation methods, over half of these did not provide adequate detail about the measures used and/or did not use a reproducible standardized measure. Of the remaining 16 different studies, there were 15 different types of scales used by patients to rate their own voice quality and VE.[13] This methodological heterogeneity makes it difficult to compare results or make recommendations for standardizing clinical methods for measuring VE in ADLD.[22,23]

Measures of vocal effort

Various methods have been proposed to measure VE in speakers with voice disorders, including ADLD, in clinical and research settings. These methods include electroglottography,[24] aerodynamic,[25] acoustic,[3] and several patient-reported measures.[11,22,23] Although VE has commonly been used to describe physiological conditions in voice production, VE is in fact a perceptual phenomenon that refers to *perceived physical exertion of the speaker during vocal tasks*.[12] In other words, VE is by its very nature a patient-perceived construct, which should therefore be measured using self-report.[12]

In addition to differences in terminology used to describe VE, there is also substantial variability in the methods and scale types used to obtain patient-perceived VE measures across settings.[23] Although no standardized clinical scale for measuring VE has been established,[22,23] studies have included interval and Likert-type scales, direct magnitude estimation scales,[26,27] the BORG CR10 scale,[23,28] and visual analog scale (VAS)s[10,29–32] for measuring VE across populations. This variability in scale types also makes it difficult to compare results across studies.

Many researchers have used VASs to measure self-rated VE in various patient populations, including ADLD.[10,29–32] A VAS is a continuous line with two defined endpoints, commonly labeled as either "no effort" or "minimum effort" versus "extreme effort" or "maximum effort" when used to measure VE.[23] A VAS provides high resolution and might be more sensitive to changes following treatment compared with scales that have fewer scale points or ranked categories.[33] The psychometric properties of VASs also may allow for measuring linear and nonlinear perceptual phenomena, although this property has not been established for VE.

Previously, Eadie et al[10] included VASs for ratings of VE and overall severity of voice in 20 speakers with ADLD. Speakers read the first paragraph of the Rainbow Passage[34] and then immediately provided self-ratings of voice quality and VE after reviewing their speech recordings. Inexperienced listeners and experienced speech-language pathologists (SLPs) also made auditory-perceptual judgments of overall severity of voice and perceived effort of the speakers' voice production using similar scales for comparison. The results showed that patient-perceived measures of voice quality and VE moderately correlated with other listeners' perceptions of similar dimensions (ranging from r = 0.45 to 0.58). The speakers' perceptions of their own VE also were more strongly correlated (r = 0.61) with Voice Handicap Index (VHI)[35] scores than other listener ratings (ranging from r = 0.30 to 0.32), consistent with the notion that other listeners' ratings of voice quality or VE are measuring different constructs than patient-reported V-RQOL. In addition, the results suggested that patient-perceived VE was a meaningful symptom in speakers with ADLD that may not be captured entirely by patient-reported quality of life measures like the VHI.

More recently, researchers have sought to adapt and validate the OMNI scales for measuring VE in individuals with voice disorders.[11] The OMNI (originating from the term, *omnibus*) scales are a series of validated scales originally developed to measure perceived exertion for different types of physical resistance exercises.[36] The OMNI uses the following operational definition of perceived exertion: "The perception of physical exertion is defined as the subjective intensity of effort, strain, discomfort, and/or fatigue" (p. 336).[37] The OMNI includes a 11-point equal-appearing interval rating scale ranging from 0 to 10 (0 = extremely easy, 2 = easy, 4 = somewhat easy, 6 = somewhat hard, 8 = hard, 10 = extremely hard). It is administered immediately following a physical task/set. The OMNI scale was adapted and validated for measuring VE in patients with ADLD.[37] The OMNI Vocal Effort Scale (OMNI-VES) includes the same operational definition of voice-related perceived exertion as the OMNI, although speakers are instructed to rate perceived VE "when using your voice TODAY" (Figure 1).[11]

In the validation study by Shoffel-Havakuk et al, 178 patients with ADLD completed several voice assessment measures directly before receiving regular BTX injections.[11] Specifically, patients completed the V-RQOL[15] and the OMNI-VES *before* Consensus Auditory Perceptual Evaluation-Voice (CAPE-V) ratings were obtained from two SLPs based on speech recordings that included a variety of tasks. A subset of 17 speakers with ADLD completed the protocol in the same order, 1 month following BTX (ie, when voice is typically improved).[11] Finally, the OMNI-VES was administered to 48 control speakers without voice disorders.

The results of the study showed that OMNI-VES scores obtained from speakers with ADLD were significantly higher, on average, than controls.[11] There also were statistically significant changes in perceived VE following BTX, demonstrating sensitivity of the scale to treatment.[11] However, OMNI-VES scores from speakers with ADLD were significantly, but only weakly related to V-RQOL scores ($\tau - b = -0.25$).[11] Finally, OMNI-VES scores from speakers were unrelated to clinicians' auditory-perceptual measures of dysphonia severity ($\tau - b = 0.08$).[11] The results led the authors to conclude that the OMNI-VES was measuring a different, but complementary, construct than other patient-reported as well as clinician-rated voice measures.

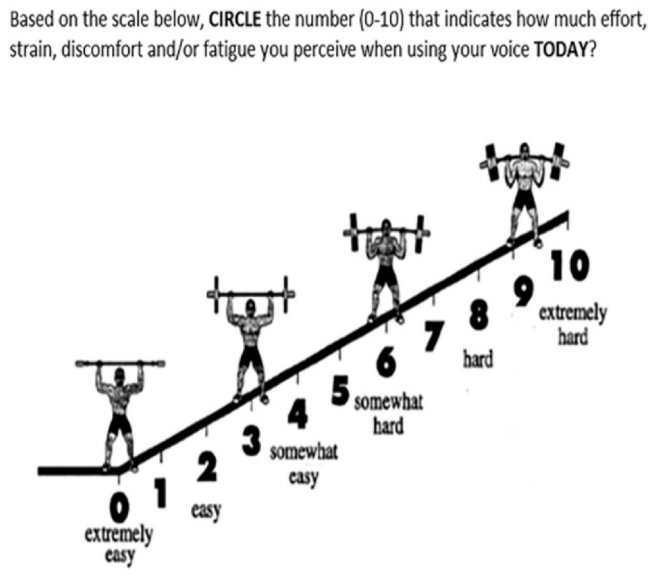

FIGURE 1. The OMNI Vocal Effort Scale.

Recently, the OMNI-VES was used in two different studies to establish normative data in control speakers,[38] as well as the effects of BTX in 26 patients with ADLD.[39] Like the initial validation study,[11] Dwyer et al found statistically significant reduction in OMNI-VES scores following BTX for patients with ADLD.[39] They also found that changes in OMNI-VES scores were most strongly related to changes in other patient-perceived measures of VE or self-rated severity of voice breaks, using VASs (rs > 0.68). Consistent with the OMNI-VES validation study and prior results by Shoffel-Havakuk et al,[11] relatively weak, nonsignificant relationships were found between changes in perceived VE and changes in V-RQOL and communicative participation from pre BTX to post BTX. However, like the original OMNI-VES validation study, Dwyer et al also reported weak relationships between changes in OMNI-VES scores and changes in clinician-rated overall severity of voice and strain (rs < 0.25).[39]

One other notable difference among these studies was the order in which perceived VE was measured as part of the voice assessment protocol. The order and timing for rating perceived VE are important when attempting to control factors that have previously been shown to affect a speaker's VE ratings, such as mood, attention, and memory.[23] In fact, authors of a consensus paper on VE recommended that because perceived VE is an in-the-moment phenomenon, such ratings "should be made immediately following a vocal activity" (p. 517).[12] In the original OMNI-VES validation study, patient-perceived measures were obtained *before* the standardized speech tasks.[11] In the study by Robertson et al,[37] patient-reported measures were completed *after* the speech tasks within a 72-hour window, although some participants did not complete all measures. In contrast, ADLD speakers in the study by Eadie et al[10] rated perceived VE immediately after reviewing their own speech recordings, anchoring VE ratings to the voice task. As such, the current study's main purpose was to examine the relationship between perceived VE rated on a VAS and the OMNI-VES when these measures were obtained directly after a vocal activity in speakers with ADLD. A second purpose was to evaluate relationships between both measures of perceived VE with other patient- and clinician-rated voice assessment measures in this population.

METHODS

All procedures were approved by the institutional review boards at Boston University and the University of Washington.

Speakers

Fifty-three individuals diagnosed with ADLD (8 M, 45 F) with or without tremor participated. All participants identified as cisgender. The mean age of speakers with ADLD was 61 years (standard deviations [SD] = 13.40 years). Forty-nine (92%) participants self-identified as White, 3 (6%) as Asian and multiple races, and 1 (2%) preferred not to answer. Forty-seven (89%) participants did not identify as Hispanic or Latino, 2 (4%) identified as Hispanic or Latino, and 4 (7%) preferred not to answer.

Fifty individuals (94%) received regular BTX injections, with the majority (n = 40, 80%) of those receiving ongoing treatment at the University of Washington Medical Center (UWMC). Individuals who received regular BTX injections at UWMC reported receiving their injections for 11.68 years on average (SD = 9.48 years) and reported responding favorably to at least two previous BTX injections. Three speakers with ADLD were not receiving current BTX treatment; of these, two reported receiving BTX previously, but had since opted to discontinue treatment. No speakers were receiving voice therapy for the reduction of symptoms associated with accompanying muscle tension dysphonia. All speakers received their ADLD diagnosis by a fellowship-trained laryngologist and experienced speech-language pathologist team based on case history, auditory-perceptual evaluation of voice, and videolaryngostroboscopic evaluation. All individuals were symptomatic, via clinician- and self-report, at the time they participated in the study. Exclusion criteria included patients < 18 years, patients who could not complete patient-reported measures in English, and patients who had been diagnosed with other speech or voice disorders.

Listeners

Ten speech-language pathologists (one M, eight F, eight cisgender, and one nonbinary) with clinical experience in voice assessment and treatment participated as listeners. Listeners were, on average, 36 years of age (SD = 8.10 years), with 9 (90%) self-identifying as White, 1 (10%) as Asian, and none identifying as Hispanic or

Latino. Listeners reported 10.45 years of clinical experience with voice disorders, on average (SD = 7.64 years, range 3-24 years). All reported experience using auditory-perceptual scales, including VAS, for regular voice assessment, and none reported difficulties with hearing.

Data collection in speakers with ADLD

Speakers with ADLD completed a variety of tasks in this study, in the following order: 1) the V-RQOL scale; 2) voice recordings; 3) patient-perceived VE using a VAS (VE-VAS) and a self-rated measure of overall severity of voice quality using a VAS (OS-VAS); and 4) patient-perceived VE using the OMNI-VES. Speakers with ADLD completed patient-reported measures using either REDCap[40,41] electronic data capture tools hosted at the University of Washington, or paper for data collection. Forty-five out of 50 (90%) participants who were receiving regular BTX injections were recorded at the end of their BTX treatment cycle, before their next injection (ie, same day). Five individuals who were receiving regular BTX injections provided assessment measures as part of a larger research protocol at Boston University and were symptomatic at the time of participation, with planned BTX injections within less than a month of their participation date.

Voice-related quality of life

Speakers with ADLD first completed the V-RQOL.[15] The V-RQOL is a 10-item, validated V-RQOL assessment tool commonly used in clinical and research settings. It includes four social-emotional domain items and six physical functioning domain items. The instructions ask patients to consider how severe and how frequent their voice problems have been during the most recent 2-week period. A 5-point Likert scale is used for measurement. Raw scores are converted to standardized totals, with 0 indicating a very poor V-RQOL, and 100 an excellent V-RQOL. For this study's purpose, both the total and physical functioning domain scores were calculated, as it was hypothesized that the physical domain items would be more strongly related to perceived VE than social-emotional items.

Voice recordings

After completing the V-RQOL, voice samples were collected from speakers with ADLD using a headset condenser microphone (AKG C420 or Shure omnidirectional MX153) routed to a Zoom H6 audio recorder or the Kay Pentax Computerized Speech Lab Model 4300 (Pine Brook, NJ), with a constant mouth-to-mic distance and microphone placed 45 degrees from midline. All recordings were obtained at a 44.1-kHz sampling rate and 16-bit resolution. All speakers performed sentence reading (eg, CAPE-V sentences, sentences loaded with voiced and voiceless phonemes), whispering, sustained phonation, counting, laryngeal diadochokinetic tasks, soft sustained voice production, and a 30-second conversational speech sample as part of a larger ongoing research protocol. Finally, speakers were asked to read the first paragraph of the Rainbow Passage.[34]

Perceived VE and self-rated overall severity of voice quality using 100-mm VAS

Before reading the Rainbow Passage and making voice recordings, speakers with ADLD were instructed that they would rate their perceived VE following the reading task, with perceived VE defined as the physical exertion in producing voice. Specifically, when using the VAS, speakers were asked to anchor their ratings to the vocal activity (ie, the standard reading passage). Upon completing the voice recording, speakers with ADLD rated their perceived VE using a 100-mm VAS (VE-VAS) (Figure 2).

Speakers also were provided an operational definition of overall severity of voice quality as "a measure of how 'good' or 'poor' the voice quality is (how clear your voice is)." They self-rated their own overall severity of voice quality using a horizontally oriented 100-mm VAS (ADLD-OS). Endpoints on the VAS were labeled 0 = normal/typical for age and gender at the extreme left and 100 = severe at the extreme right (Figure 2).

Perceived VE using the OMNI-VES

Finally, speakers with ADLD read the standard instructions and definitions of perceived VE using the OMNI-VES.[11] As found in the current validated version of the OMNI-VES, they were instructed to rate their perception of VE "when using your voice TODAY" (Figure 1).

Speech stimuli preparation and listener procedures

At the time the listening procedure was conducted, data were available for a subset of participants (39/53, or 74% of the speech samples). ADLD and vocal tremor were reported in 8/39 (21%) of these speakers and verified as part of their medical record. The second sentence of the Rainbow Passage was extracted from the available subset of voice recordings. Speakers performed the same speech tasks, and were either at the end of their BTX cycles (or did not undergo regular BTX injections). The samples were peak-normalized for the listening experiment using sound-editing software. Samples were then entered into a custom software program that randomly generates speaker order, presents perceptual rating scales, and records responses.

Ten experienced SLP listeners were then asked to perform auditory-perceptual ratings of the speech samples. Listeners were first oriented to the rating task and provided an operational definition of overall severity, a measure of how good or poor the voice sample is judged to be for its voice quality. Listeners were then presented four speech samples (two male and two female speakers with ADLD) that represented a range of voice severities, from normal/typical to severe, for familiarization purposes. None of these speakers were included in the subsequent rating task.

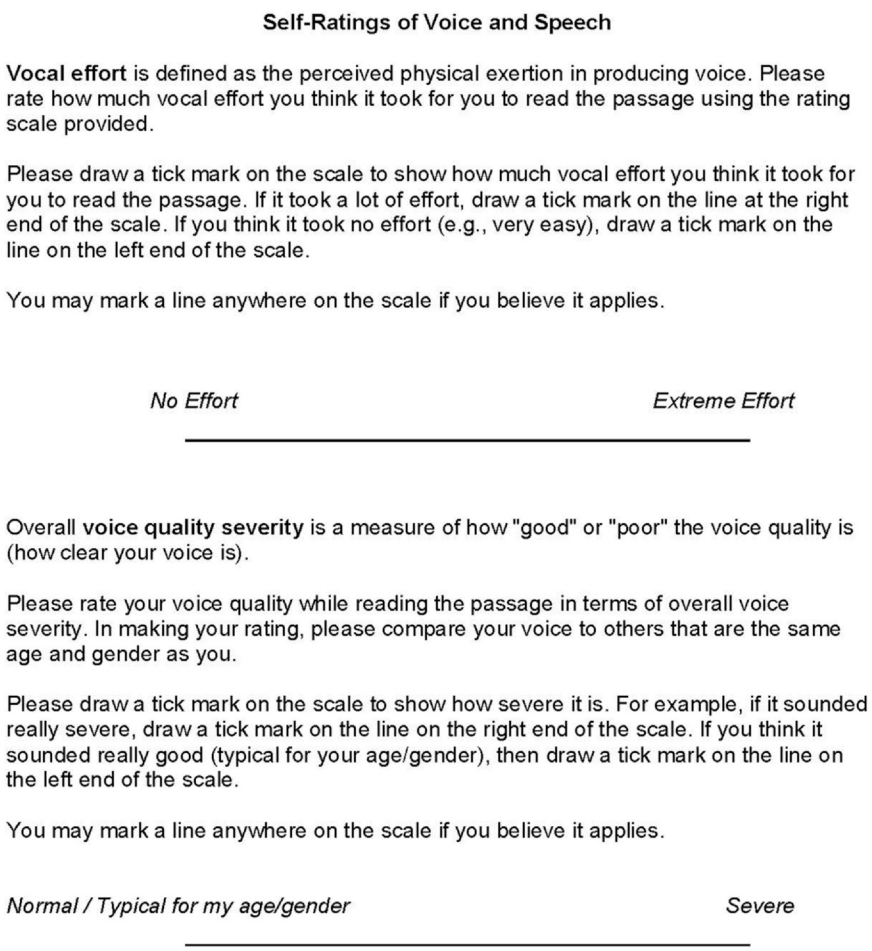

FIGURE 2. The Vocal Effort Visual Analog Scales and Voice Quality Severity Scales used for self-ratings.

Listeners then provided auditory-perceptual ratings of 39 speakers with ADLD for overall severity of voice (SLP-OS) using a horizontally oriented 100-mm VAS, similar to the CAPE-V.[9] Endpoints on the VAS were labeled 0 = typical at the extreme left and 100 = severe at the extreme right. Stimuli were presented in random order via a custom-made computer program over Sony MDR7506 headphones. Listeners could listen to each sample as many times as they wished. All 39 voice samples were repeated in random order and rated at least 7 days after the initial rating task to assess intrarater reliability.

Statistical analyses

A power analysis was conducted to calculate the sample sizes needed to detect a relevant simple correlation ($r = 0.4$) using a 5% significance level ($\alpha = 0.05$) with 80% power ($\beta = 0.20$) using a two-sided test. The required sample size was approximately 47 (N = 47). The strength of the relationship between the two patient-perceived VE measures, VE-VAS and the OMNI-VES, was evaluated using Spearman's correlation coefficient, a nonparametric statistic, because data were not normally distributed. Correlational analyses were also calculated to evaluate the strength of relationships between both measures of VE and other patient-rated voice assessment measures, including self-rated overall voice quality (ADLD-OS) and the V-RQOL total and physical functioning domain scores. Correlational analyses were also calculated to determine the strength of relationships between both measures of VE and clinician auditory-perceptual ratings of overall voice severity (SLP-OS). Finally, simple linear regression was conducted to determine the proportion of variance in VE ratings that could be explained by patient-rated and clinician-rated measures.

Reliability for clinician-rated overall severity of voice (SLP-OS)

To calculate intrarater reliability, listeners repeated the auditory-perceptual task in a second rating session held at least 7 days after the first session (mean time between sessions 1 and 2 = 17.4 days). Intrarater reliability was calculated using an individual listener's first and second ratings of all repeated stimuli using Spearman's correlation coefficient because data were not normally distributed. Interrater reliability for listeners' judgments was calculated using intraclass correlations (ICCs) based on a single rating

TABLE 1.
Means, Standard Deviations, and Ranges of Scores for Patient- and Clinician-Rated Outcome Measures

Construct Measured	Patient- or Clinician-Rated	Scale Name (Abbr)	N	Mean	(SD)	Range
Perceived VE	Patient	VE-VAS	52	60.27 mm	(22.18)	12-99 mm
Perceived VE	Patient	OMNI-VES	52	6.77	(1.54)	4-10
Overall severity of voice quality	Patient	ADLD-OS	52	68.65 mm	(21.38)	15-100 mm
Voice-related quality of life	Patient	V-RQOL total	52	51.60	(17.33)	13-88
Voice-related quality of life	Patient	V-RQOL—physical functioning domain	50	45.76	(21.10)	2-88
Overall severity of voice quality	Clinician	SLP-OS	39	47.91 mm	(27.51)	4-94 mm

(k = 1), absolute-agreement, two-way random-effects model for ratings from session 1.

The average intrarater reliability of overall severity of voice quality ratings across all 10 listeners (OS-SLP) from session 1 to session 2 was good: mean $r = 0.88$.[42] Interrater reliability was also good: ICC(2, 1) = 0.85, 95% CI [0.78, 0.91]. As a result, listener ratings were deemed adequate for interpretation and further statistical analyses.

RESULTS

Descriptive statistics

The means, SD, and range of scores for patient-rated measures of VE, overall severity of voice quality (ADLD-OS), and V-RQOL total and physical functioning domain scores obtained from speakers with ADLD are presented in Table 1. Means, standard deviation, and ranges also are presented for 10 experienced listeners' ratings of overall severity of voice for 39 speakers with ADLD.

The mean VE-VAS rated by speakers with ADLD (n = 52) was 60.27 mm (SD = 22.18 mm), consistent with a moderate-to-severe perceived VE. The mean OMNI-VES scores for speakers with ADLD (n = 52) were 6.77 (SD = 1.54), consistent with a perception of "somewhat hard" VE. Ratings of overall severity of voice quality also were obtained from 52 speakers with ADLD (ADLD-OS). The mean ADLD-OS rating was 68.65 mm (SD = 21.38 mm), consistent with the perception of severe dysphonia.[33] Speakers with ADLD also reported mean standardized total scores of 51.60 (SD = 17.33) and mean physical functioning domain scores of 45.76 (SD = 21.10) on the V-RQOL, demonstrating moderate V-RQOL.[15] Finally, SLP listeners rated the subset of 39 speakers with ADLD as having a mean overall voice quality severity rating of 47.91 mm (SD = 27.51 mm).

Relationship between measures of VE

There was a strong,[42] statistically significant correlation ($r = 0.78$, $P < 0.001$) between patient-perceived VE rated on the VAS (VE-VAS) and scores on the OMNI-VES. The line of best fit and the predicted variance score ($r^2 = 0.67$) also demonstrate the strong linear relationship between the two measures of VE (Figure 3).

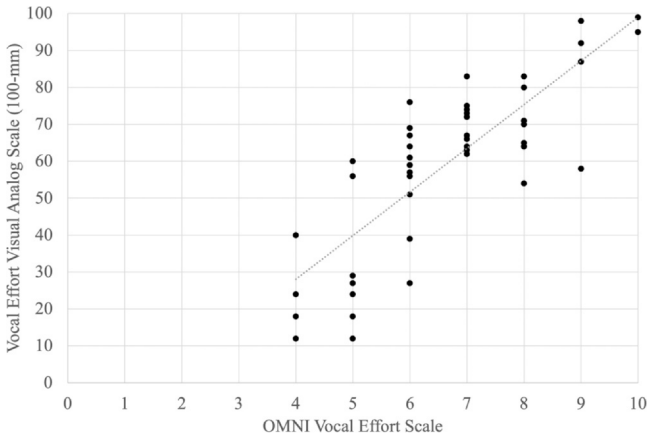

FIGURE 3. Correlation between ratings of vocal effort using a 100-mm visual analog scale and the OMNI Vocal Effort Scale.

Relationships between VE and other patient-rated measures

Ratings of perceived VE on the 100-mm VAS (VE-VAS) were strongly[42] and significantly related to self-rated severity of voice quality (VE-VAS vs ADLD-OS; $r = 0.75$, $P < 0.001$). Similarly, patient-reported OMNI-VES scores were strongly[42] and significantly correlated with self-rated severity of voice quality ($r = 0.85$, $P < 0.001$). The lines of best fit and the predicted variance scores for the VE-VAS ($r^2 = 0.65$) and for the OMNI-VES ($r^2 = 0.64$) demonstrated strong linear relationships between self-perception of voice quality and both measures of VE.

In contrast, a weak correlation[42] was found between patient-perceived VE on the VAS (VE-VAS) and standardized V-RQOL total scores ($r = -0.21$, $P > 0.05$). The line of best fit and the predicted variance score ($r^2 = 0.07$) also demonstrated the weak linear relationship between VE-VAS and V-RQOL. A similarly weak[42] correlation was found between the VE-VAS and the physical functioning domain scale of the V-RQOL ($r = -0.19$, $P > 0.05$). A weak, but significant, correlation was found between the OMNI-VES and V-RQOL total scores ($r = -0.37$, $P = 0.008$). The line of best fit and the predicted variance score ($r^2 = 0.13$) demonstrated the weak linear relationship between patient-perceived VE and V-RQOL. A more moderate[42] and significant correlation between the OMNI-

VES and the physical functioning domain scale of the V-RQOL ($r = -0.44$, $P = 0.002$) was observed.

Relationships between VE and listeners' perceptions of overall severity of voice quality

VE scores were moderately[42] and significantly related to listeners' perceptions of overall severity of voice quality with both VE measures demonstrating similar relationships with clinicians' perceptual judgments (VE-VAS vs SLP-OS; $r = 0.50$, $P < 0.001$; OMNI-VES vs SLP-OS, $r = 0.57$, $P < 0.001$). That is, increases in perceived VE by speakers with ADLD were associated with clinicians' perceptions of increased voice severity. The lines of best fit and the predicted variance scores for VE-VAS ($r^2 = 0.32$) and OMNI-VES ($r^2 = 0.35$) also demonstrated weaker linear relationships between these measures compared with speakers' self-ratings of voice quality.

DISCUSSION

VE is an important perceptual construct experienced by individuals with ADLD during speech production.[10,11,20,35] It also is a meaningful treatment outcome measure for individuals with ADLD who receive regular BTX injections.[9] Many different types of scales have been used to measure perceived VE in ADLD, making it difficult to compare results across studies. In addition, recent research groups have recommended that perceived VE be measured directly following a vocal activity because factors such as attention and memory may affect perceptions of VE. Consequently, this study examined the relationship between perceived VE rated on a VAS and the OMNI-VES when these measures were obtained directly after a vocal activity. We also evaluated relationships between both measures of VE with other patient- and clinician-rated voice measures to permit comparison with existing literature.

In this study, speakers with ADLD, on average, rated their VE as moderate to severe, rated their overall voice quality as poor, and reported a moderate V-RQOL at the time their data were obtained. Fifty out of 53 (94%) individuals regularly received BTX injections, and like many previous studies, their data were collected at the end (or near the end) of their regular BTX cycles when they were most symptomatic. The average perceived VE score was 60-mm, with speakers using a large range of the scale (range 12-99 mm). These VE ratings are similar to those reported by Dwyer et al[39] for ADLD speakers whose data were also collected just before their next BTX injections, and who rated VE for face-to-face and telephone conversations using 100-mm VAS (M = 54-mm conversation, M = 75-mm telephone). However, although speakers in the current study engaged in a vocal activity and were asked to anchor VE ratings on the VAS to that activity, it is not known how long it had been since speakers in the study by Dwyer et al[39] had engaged in a conversation face-to-face or on the telephone at the time of data collection.

The average OMNI-VES score for speakers in this study was 6.77 (SD = 1.54), consistent with a perception of "somewhat hard" VE (range 4-10). These scores are similar to OMNI-VES scores (6.7 ± 1.6, range = 4-9) reported by Dwyer et al.[39] Average OMNI-VES scores in the current study were slightly higher than the average for 178 ADLD speakers (5.07 ± 2.18) in the original OMNI-VES validation study by Shoffel-Havakuk et al,[11] in which speakers completed the OMNI-VES *before* their voice recordings. Despite methodological differences between studies that included the OMNI-VES, the average OMNI-VES scores reported by speakers with ADLD in this study were increased compared with controls without voice disorders in two previous studies, with no speakers with ADLD rating VE as < 4 in this study.[11,38]

Most importantly, when speakers with ADLD in the present study completed ratings of VE on two scales after a vocal activity, their ratings of VE were strongly and significantly correlated ($r = 0.78$), despite using different scales and instructions for rating perceived VE (100-mm VAS vs OMNI-VES). When using the VAS, speakers were asked to anchor their ratings to the vocal activity (ie, a standard reading passage). In contrast, speakers with ADLD in this study completed the OMNI-VES using standard instructions to rate their "voice as it is today." While it is unknown whether speakers in this study used the same vocal task as a referent for making OMNI-VES ratings as they did when they used the VAS, it is noteworthy that ratings were strongly correlated after speakers performed data collection in this order. Whether OMNI-VES scores would change in the same set of speakers before and after a vocal activity or in response to varied speech tasks that elicit increased symptoms in speakers with ADLD was not measured and needs future study.

In patients with voice disorders, increased VE might relate to physiological mechanisms (eg, extrinsic laryngeal tension, subglottal pressure), viscoelastic properties of the vocal folds, and cognitive-emotional factors.[23,30,43–45] For example, the task-specific involuntary intrinsic laryngeal adductor contractions and subsequent elevated subglottic pressure might contribute to increased VE in speakers with ADLD.[13,31,46] Increased supraglottic compression might also result from an attempt to compensate for these involuntary movements and contribute to VE.[32,47,48] There is also some evidence to suggest that laryngeal sensory feedback alterations might also be present in speakers with ADLD.[49] Finally, because factors such as mood, attention, and memory have been shown to influence perceived VE, it is important to investigate how order of evaluation tasks might affect VE. How perceived VE relates to other patient- and clinician-rated voice measures may help elucidate the importance of these effects on VE.[13,31,46]

Relationship between VE and other patient-rated measures

In addition to completing two measures of patient-perceived VE, speakers with ADLD in this study rated their own

overall severity of voice quality and completed the V-RQOL. The results first showed that both VE measures were strongly related to speakers' ratings of their severity of voice quality ($rs = 0.75$-0.85), regardless of the scale used to measure VE. The strength of the relationship between patient-perceived VE measures and their overall severity of voice quality in this study was slightly stronger compared with findings from Eadie et al[10] ($r = 0.67$), who similarly used a VAS to capture VE. Dwyer et al[39] also reported a moderate-to-strong relationship between ADLD patients' self-rated severity of voice breaks, as rated on a VAS, and OMNI-VES scores ($\rho = 0.58$-0.82) pre BTX and post BTX, respectively. Statistically significant correlations between changes in self-rated severity of voice breaks and changes in OMNI-VES scores post BTX were also observed ($\rho = 0.69$). The relatively strong relationships between speakers' perceptions of VE and self-rated severity of voice quality are not surprising, given that speakers with ADLD qualitatively report increased VE with increased severity of dysphonia.[9] Yet, although related, the bases of these percepts may be somewhat different, for example, speakers have access to other kinesthetic and physiologic information for making judgments of VE that go beyond the acoustic signal.[50]

Several patient-reported outcome measures, including the V-RQOL and the VHI, were designed to measure V-RQOL: a multidimensional construct that may include physical, social, emotional, and functional factors, among others.[15,35] For example, the V-RQOL captures the physical functioning and social-emotional consequences of a voice problem.[35] Speakers with ADLD in this study, on average, demonstrated moderate V-RQOL, with average V-RQOL scores (51.60 ± 17.33) generally similar to the average scores reported by Shoffel-Havakuk et al[11] (46.97 ± 29.21) and Hogikyan et al[18] (42.32 ± 22.99) when these scores were obtained just before repeated BTX injection. Further, results from this study showed that relationships between measures of VE and V-RQOL, measured on the V-RQOL, were weak ($rs = -0.21$ to -0.37), regardless of the scale used to measure VE. Shoffel-Havakuk et al[11] similarly found a weak correlation between OMNI-VES scores and V-RQOL scores ($\tau - b = -0.25$).

In this study, the relationships between the two measures of VE and V-RQOL were weaker than the relationship previously reported by Eadie et al ($r = 00.61$), who used the 30-item VHI[35] to measure V-RQOL in ADLD speakers.[10] Baldner et al also reported a moderate correlation between speakers' perceptions of VE using the Borg category-ratio 10 scale, and V-RQOL using the VHI ($r = 0.54$).[23] Their study included speakers with various types of voice disorders, including laryngeal dystonia. This contrast in findings might relate, in part, to use of different V-RQOL tools. For example, the VHI (30-item version) includes several items in the physical domain that may more closely relate to VE, including items that are designed to measure a speaker's perception of strain as well as their effort to speak. However, the questions included in the V-RQOL do not directly address VE; this is one possible reason somewhat stronger correlations might be expected between speakers' perceptions of VE and V-RQOL in studies that use the VHI (30-item version) versus the V-RQOL.[15] Still, others have failed to find consistently strong relationships between perceived VE and V-RQOL, even when using the single item on the VHI that targets it, suggesting that other factors influence ratings on quality of life questionnaires.[51] This is likely because VE measures and V-RQOL instruments measure different constructs. As complementary but important voice measures, clinicians should continue to include both types of measures when evaluating speakers with ADLD.[11,22,23]

Relationships between VE and listeners' perceptions of overall severity of voice quality

In this study, experienced SLP listeners, on average, rated speakers with ADLD as being moderately dysphonic[33] ($M = 47.91$ mm) using a 100-mm VAS. The ratings in this study were similar to experienced listeners' perceptions of overall severity of voice quality in speakers with ADLD reported by Eadie et al[10] and Dwyer et al[39] ($M = 47.2$ mm), before BTX. Interestingly, in this study, both patient-perceived measures of VE were moderately related to clinicians' perceptions of overall severity of voice quality ($rs = 0.50$-0.57) when both VE measures were obtained after a vocal activity. These results contrast with those reported in the original OMNI-VES validation study ($\tau - b = 0.08$), and those by Dwyer et al, who reported weak relationships between changes in OMNI-VES scores and changes in clinician-rated overall severity of voice and strain ($rs < 0.25$).[39]

The strength of relationships between patient-perceived VE and clinician-rated voice quality in the two studies using the OMNI-VES to measure VE[11,39] in ADLD was reduced compared with findings from the study by Eadie et al[10] that used a VAS to measure VE. As such, variability in these relationships' strength might relate to differences in the scales used for measuring VE (OMNI-VES vs VAS). One other notable difference among these studies was the order in which perceived VE was measured as part of the voice assessment protocol. In the original OMNI-VES validation study, patient-perceived measures were obtained *before* the standardized speech tasks.[11] In the study by Dwyer et al,[39] patient-reported measures were completed *after* the speech tasks. However, some participants did not complete all measures, and participants were allowed to complete patient-reported measures within a 72-hour window. In contrast, ADLD speakers in the study by Eadie et al[10] rated perceived VE immediately after reviewing their own speech recordings, and were explicitly instructed to anchor VE ratings to the voice task.

Likewise, in the present study, speakers with ADLD completed voice recordings and then were instructed to anchor VE ratings on the VAS to voice production during the voice task. Although speakers in this study completed the OMNI-VES according to standard instructions, their OMNI-VES scores were also moderately related to clinician's

perceptions of overall severity of voice quality. These results appear to support the contention that order of evaluation tasks may impact OMNI-VES scores and/or affect the strength of the relationship between OMNI-VES and clinicians' perceptions of a speaker's overall voice severity, particularly when both the speakers' and listeners' ratings may have a similar referent. Findings from this study add to the existing literature and strengthen the validity of VE measures in speakers with ADLD. However, there are additional sources of variability in VE measures that were not evaluated in this study that warrant further investigation.

LIMITATIONS AND FUTURE DIRECTIONS

There are several limitations of this study. One of these limitations relates to heterogeneity of the speakers with ADLD. Due to sample size limitations, we were unable to evaluate the effect of time since symptom onset, presence of concomitant dystonic or essential vocal tremor, age, response to BTX, or other factors on the results. In this study, the OMNI-VES and VE-VAS were highly correlated and the strength of the relationship between these both VE measures and auditory-perceptual measures was similar. However, we did not control for a possible order effect. That is, patient-perceived VE ratings on the VAS were always obtained before the OMNI-VES. Speakers rated the amount of VE required to produce voice during the reading passage when using the VAS, whereas the OMNI-VES instructed speakers to rate their perceived VE when using their voice today. Hence, it is unknown whether speakers rated VE as an "in the moment phenomena"[52] when using the OMNI-VES rating scale, despite obtaining both VE measures immediately after the same voice tasks. A second major limitation relates to the interpretation of the results; relationships were assessed using correlations, and therefore, readers should be cautioned not to interpret a change in one measure as causing a change in another measure.

Clinicians and researchers should be aware of potential task and order effects when measuring perceived VE in ADLD, regardless of the scale used to measure perceived VE. Because factors such as memory, attention, or mood may affect perceived VE, researchers recently recommended that such ratings "be made immediately following a vocal activity" (p. 517).[12] This recommendation aligns with procedures used in the original OMNI scales that measure perceived exertion directly following a physical task/set. Whether instructions on the OMNI-VES, in particular, need revision should be a subject of further research. In addition, future studies should continue to investigate whether there are advantages of one type of VE scale over another (eg, psychometric properties, patient preference, efficiency of measure, etc) when measuring patient-perceived VE in ADLD.

CONCLUSION

Speakers with ADLD typically present with increased VE, increased dysphonia, and moderate V-RQOL at the end of their regular BTX cycle. However, although improved VE is one important measure of treatment success, standardized clinical measures of VE are lacking.[13,22] When patient-reported measures of VE are performed directly after a vocal activity, they are moderately and strongly related to clinicians' and speakers' auditory-perceptual voice ratings, respectively. Measures of VE are weakly related to V-RQOL measures, similar to results from prior studies. These findings contribute to our understanding of the convergent and discriminant validity of measures of VE. The results are also useful for considering appropriate assessment protocols for speakers with ADLD during their BTX treatment cycle, as well as tracking the effects of meaningful ADLD symptoms in individuals who choose not to undergo BTX treatment.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships that may be considered as potential competing interests: Cara Sauder, Cara Stepp, Tanya Eadie, and Katherine Marks reports financial support was provided by National Institutes of Health. Tanya Eadie reports financial support was provided by Dysphonia International. The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by grant R01DC015570 from NIH/NIDCD (Stepp: PI). REDCap was also used to collect patient-reported data, supported by UL TR0002319. The senior author (Eadie) was awarded the 6th Annual Dysphonia International Research Award from the Voice Foundation and Dysphonia International based on this work. The authors thank the participants and members of the Vocal Function Lab for their contributions to this project.

References

1. Mor N, Blitzer A. Diagnosis and treatment of laryngeal dystonia: past, present and future directions. *Tremor Other Hyperkinetic Mov.* 2016:1–12. https://doi.org/10.7916/D8G160J5.
2. Roy N, Gouse M, Mauszycki SC, et al. Task specificity in adductor spasmodic dysphonia versus muscle tension dysphonia. *Laryngoscope.* 2005;115:311–316. https://doi.org/10.1097/01.mlg.0000154739.48314.ee.
3. Eadie TL, Stepp CE. Acoustic correlate of vocal effort in spasmodic dysphonia. *Ann Otol Rhinol Laryngol.* 2013;122:169–176. https://doi.org/10.1177/000348941312200305.
4. Langeveld TPM, Drost HA, Zwinderman AH, et al. Perceptual characteristics of adductor spasmodic dysphonia. *Ann Otol Rhinol Laryngol.* 2000;109:741–748. https://doi.org/10.1177/000348940010900808.
5. Ludlow C, Domangue R, Sharma D, et al. Consensus-based attributes for identifying patients with spasmodic dysphonia and other voice disorders. *JAMA Otolaryngol Head Neck Surg.* 2018;144:657–665.

6. Murry T. Spasmodic dysphonia: let's look at that again. *J Voice.* 2014;28:694–699. https://doi.org/10.1016/j.jvoice.2014.03.007.
7. Sauder CL, Kapsner-Smith MR, Simmons E, et al. The effect of rating method on reliability of judgments of strain across populations. *Am J Speech Lang Pathol.* 2024;33:393–405. https://doi.org/10.1044/2023_AJSLP-23-00174.
8. Stewart CF, Allen EL, Tureen P, et al. Adductor spasmodic dysphonia: standard evaluation of symptoms and severity. *J Voice.* 1997;11:95–103. https://doi.org/10.1016/S0892-1997(97)80029-X.
9. Baylor CR, Yorkston KM, Eadie TL. The consequences of spasmodic dysphonia on communication-related quality of life: a qualitative study of the insider's experiences. *J Commun Disord.* 2005;38:395–419. https://doi.org/10.1016/j.jcomdis.2005.03.003.
10. Eadie TL, Nicolici C, Baylor C, et al. Effect of experience on judgments of adductor spasmodic dysphonia. *Ann Otol Rhinol Laryngol.* 2007;116:695–701. https://doi.org/10.1177/000348940711600912.
11. Shoffel-Havakuk H, Marks KL, Morton M, et al. Validation of the OMNI vocal effort scale in the treatment of adductor spasmodic dysphonia: OMNI Vocal Effort Scale in ADSD. *Laryngoscope.* 2019;129:448–453. https://doi.org/10.1002/lary.27430.
12. Hunter EJ, Cantor-Cutiva LC, Van Leer E, et al. Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. *J Speech Lang Hear Res.* 2020;63:509–532. https://doi.org/10.1044/2019_JSLHR-19-00057.
13. Rumbach A, Aiken P, Novakovic D. Treatment outcome measures for spasmodic dysphonia: a systematic review. *J Voice.* 2024;38:540.e13–540.e43. https://doi.org/10.1016/j.jvoice.2021.10.001.
14. Bielamowicz S, Ludlow CL. Effects of botulinum toxin on pathophysiology in spasmodic dysphonia. *Ann Otol Rhinol Laryngol.* 2000;109:194–203. https://doi.org/10.1177/000348940010900215.
15. Hogikyan ND, Hogikyan ND, Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *J.* 1999;13:557–569.
16. Rubin AD, Wodchis WP, Spak C, et al. Longitudinal effects of botox injections on voice-related quality of life (V-RQOL) for patients with adductory spasmodic dysphonia: part II. *Arch Otolaryngol Neck Surg.* 2004;130:415. https://doi.org/10.1001/archotol.130.4.415.
17. Yiu Y, Baylor CR, Bamer AM, et al. Validation of the communicative participation item bank as an outcome measure for spasmodic dysphonia. *Laryngoscope.* 2021;131:859–864. https://doi.org/10.1002/lary.28897.
18. Hogikyan ND, Wodchis WP, Spak C, et al. Longitudinal effects of botulinum toxin injections on voice-related quality of life (V-RQOL) for patients with adductory spasmodic dysphonia. *J Voice.* 2001;15:576–586. https://doi.org/10.1016/S0892-1997(01)00060-1.
19. Chang CY, Chabot P, Thomas JP. Relationship of botulinum dosage to duration of side effects and normal voice in adductor spasmodic dysphonia. *Otolaryngol Neck Surg.* 2007;136:894–899. https://doi.org/10.1016/j.otohns.2007.01.035.
20. Braden MN, Johns MM, Klein AM, et al. Assessing the effectiveness of botulinum toxin injections for adductor spasmodic dysphonia: clinician and patient perception. *J Voice.* 2010;24:242–249. https://doi.org/10.1016/j.jvoice.2008.08.003.
21. Kreiman J, Gerratt BR. Validity of rating scale measures of voice quality. *J Acoust Soc Am.* 1998;104:1598–1608. https://doi.org/10.1121/1.424372.
22. Van Mersbergen M, Beckham BH, Hunter EJ. Do we need a measure of vocal effort? Clinician's report of vocal effort in voice patients. *Perspect ASHA Spec Interest Groups.* 2021;6:69–79. https://doi.org/10.1044/2020_PERSP-20-00258.
23. Ford Baldner E, Doll E, Van Mersbergen MR. A review of measures of vocal effort with a preliminary study on the establishment of a vocal effort measure. *J Voice.* 2015;29:530–541. https://doi.org/10.1016/j.jvoice.2014.08.017.
24. Somanath K, Mau T. A measure of the auditory-perceptual quality of strain from electroglottographic analysis of continuous dysphonic speech: application to adductor spasmodic dysphonia. *J Voice.* 2016;30:770.e9–770.e21. https://doi.org/10.1016/j.jvoice.2015.11.005.
25. Witsell DL, Weissler MC, Martinkosky SJ, et al. Measurement of laryngeal resistance in the evaluation of botulinum toxin injection for treatment of focal laryngeal dystonia. *Laryngoscope.* 1994;104:8–11. https://doi.org/10.1288/00005537-199401000-00003.
26. Verdolini K, Titze IR, Fennell A. Dependence of phonatory effort on hydration level. *J Speech Hear Res.* 1994;37:1001–1007. https://doi.org/10.1044/jshr.3705.1001.
27. McHenry M, Evans J, Powitzky E. Singers' phonation threshold pressure and ratings of self-perceived effort on vocal tasks. *J Voice.* 2013;27:295–298. https://doi.org/10.1016/j.jvoice.2012.12.013.
28. Van Leer E, Van Mersbergen M. Using the Borg CR10 physical exertion scale to measure patient-perceived vocal effort pre and post treatment. *J Voice.* 2017;31:389.e19–389.e25. https://doi.org/10.1016/j.jvoice.2016.09.023.
29. Eadie TL, Kapsner-Smith M. The effect of listener experience and anchors on judgments of dysphonia. *J Speech Lang Hear Res.* 2011;54:430–447. https://doi.org/10.1044/1092-4388(2010/09-0205).
30. Tanner K, Fujiki RB, Dromey C, et al. Laryngeal desiccation challenge and nebulized isotonic saline in healthy male singers and nonsingers: effects on acoustic, aerodynamic, and self-perceived effort and dryness measures. *J Voice.* 2016;30:670–676. https://doi.org/10.1016/j.jvoice.2015.08.016.
31. McCabe DJ, Titze IR. Chant therapy for treating vocal fatigue among public school teachers: a preliminary study. *Am J Speech Lang Pathol.* 2002;11:356–369. https://doi.org/10.1044/1058-0360(2002/040).
32. McKenna VS, Diaz-Cadiz ME, Shembel AC, et al. The relationship between physiological mechanisms and the self-perception of vocal effort. *J Speech Lang Hear Res.* 2019;62:815–834. https://doi.org/10.1044/2018_JSLHR-S-18-0205.
33. Karnell MP, Melton SD, Childes JM, et al. Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *J Voice.* 2007;21:576–590. https://doi.org/10.1016/j.jvoice.2006.05.001.
34. Fairbanks G. *Voice and Articulation Drillbook*. 2nd ed. New York, NY: Harper & Row; 1960.
35. Jacobson BH, Johnson A, Grywalski C, et al. The Voice Handicap Index (VHI): development and validation. *Am J Speech-Lang Pathol.* 1997;6:66.
36. Robertson RJ. Exercise testing and prescription using RPE as a criterion variable. *Int J Sport Psychol.* 2001;32:177–188.
37. Robertson RJ, Goss FL, Rutkowski J, et al. Concurrent validation of the OMNI perceived exertion scale for resistance exercise. *Med Sci Sports Exerc.* 2003;35:333–341. https://doi.org/10.1249/01.MSS.0000048831.15016.2A.
38. Morton-Jones ME, Timmons Sund L, Castro ME, et al. Adult normative data for the OMNI-Vocal Effort Scale (VES). *Laryngoscope.* 2024;134:3726–3731. https://doi.org/10.1002/lary.31464.
39. Dwyer CD, Gochman GE, Rosen CA, et al. Comparison of outcome measures (subjective, objective, and patient-based) in laryngeal dystonia treatment with botulinum toxin A injection. *J Voice.* 2023. https://doi.org/10.1016/j.jvoice.2023.03.019.In press.
40. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42:377–381. https://doi.org/10.1016/j.jbi.2008.08.010.
41. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform.* 2019;95:103208. https://doi.org/10.1016/j.jbi.2019.103208.
42. Portney L, Watkins M. *Construct validity. Foundations of Clinical Research: Applications to Practice*. Upper Saddle River, NJ: Prentice Hall Health; 2000:87–91.
43. Van Mersbergen M, Delany M. Vocal responses to emotional picture viewing. *Logoped Phoniatr Vocol.* 2014;39:99–107. https://doi.org/10.3109/14015439.2013.777108.
44. Chang A, Karnell MP. Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: a study of vocal fatigue. *J Voice.* 2004;18:454–466. https://doi.org/10.1016/j.jvoice.2004.01.004.

45. Solomon NP, Glaze LE, Arnold RR, et al. Effects of a vocally fatiguing task and systemic hydration on men's voices. *J Voice.* 2003;17:31–46. https://doi.org/10.1016/S0892-1997(03)00029-8.
46. Kempster GB, Gerratt BR, Verdolini Abbott K, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol.* 2009;18:124–132. https://doi.org/10.1044/1058-0360(2008/08-0017).
47. Simonyan K, Barkmeier-Kraemer J, Blitzer A, et al. Laryngeal dystonia: multidisciplinary update on terminology, pathophysiology, and research priorities. *Neurology.* 2021;96:989–1001. https://doi.org/10.1212/WNL.0000000000011922.
48. Mor N, Simonyan K, Blitzer A. Central voice production and pathophysiology of spasmodic dysphonia. *Laryngoscope.* 2018;128:177–183. https://doi.org/10.1002/lary.26655.
49. Young VN, Kidane J, Gochman GE, et al. Abnormal laryngopharyngeal sensation in adductor laryngeal dystonia compared to healthy controls. *Laryngoscope.* 2023;133:2271–2278. https://doi.org/10.1002/lary.30462.
50. Brandt JF, Ruder KF, Shipp T. Vocal loudness and effort in continuous speech. *J Acoust Soc Am.* 1969;46:1543–1548. https://doi.org/10.1121/1.1911899.
51. Tanner K, Roy N, Merrill RM, et al. Spasmodic dysphonia: onset, course, socioemotional effects, and treatment response. *Ann Otol Rhinol Laryngol.* 2011;120:465–473. https://doi.org/10.1177/000348941112000708.
52. Borg G. Psychophysical scaling with applications in physical work and the perception of exertion. *Scand J Work Environ Health.* 1990;16:55–58. https://doi.org/10.5271/sjweh.1815.