# A Review of Measures of Vocal Effort With a Preliminary Study on the Establishment of a Vocal Effort Measure

*,[1]Elizabeth Ford Baldner, †Emerald Doll, and *Miriam Ruth van Mersbergen, *DeKalb, Illinois and †Madison, Wisconsin

**Summary: Objectives/Hypothesis.** Vocal effort is the perceived exertion of voicing. Patients with voice disorders (VDs) frequently complain of increased effort, but currently there exists no empirically validated scale for the measurement of vocal effort.

This article reviews the extant literature on vocal effort and its various definitions. It also presents a preliminary investigation on the use of a psychophysical scale, the Borg Category Ratio (CR-10), for vocal effort ratings.

**Study Design.** A total of 28 participants with VDs and 28 healthy controls (HCs) underwent acoustic and aerodynamic voice measures in this prospective quasi-experimental group design.

**Methods.** Vocal effort ratings using the Borg CR-10 gathered on vowels, standard sentences, and conversation were correlated with auditory-perceptual ratings, Voice Handicap Index scores, and phonation threshold pressure (PTP).

**Results.** Results indicate that the Borg CR-10 is not sensitive to the presence of a VD but does correlate moderately well with other measures of VD severity, and may be clinically indicated for such use.

**Conclusions.** Future research directions include task choice for vocal effort ratings, considerations during PTP protocols, intensive examiner and examinee training, and use of the Borg CR-10 for within-group separation of VD diagnosis.

**Key Words:** Vocal effort–Scale validation–Voice–Perceived phonatory exertion.

## INTRODUCTION

For people with healthy vocal mechanisms in normal acoustic environments, speech requires negligible vocal effort. Vocal effort typically increases in response to external influences, such as overcoming loud ambient noise, projecting across a great distance, or by temporary intrusion in the case of acute illness or overuse.[1,2] However, if the vocal task becomes chronically and abnormally effortful, individuals can be aware of vocal effort impinging on communication, motivating them to seek professional treatment.[3] The voice literature is resplendent with appeals to uncover the most scientifically and clinically appropriate scale to measure vocal effort. The need to balance objective measures with subjective experience is recognized in the voice literature, and vocal effort can be the most frequently reported and most endured patient complaint[4] with a call to systematically assess the relation of perceived and physiological effort with voicing.[5] This article will discuss some of the concepts of vocal effort as they relate to the tools that purport to measure this effort.

### Parameters of effort

Effort is a perceived exertion an individual feels from within. It is characterized by the cognitive, kinesthetic, and dynamic bodily response to physical activity.[6,7] A challenging metric to measure and describe effort requires both the perceptual input of the individual along with physical data necessary for scientific evaluation.[5,6] As with all perceptual measures, capturing an internal sensation and relating it to a physical phenomenon or measurably quantifying it requires careful and diligent thought. This is particularly difficult when trying to quantify vocal effort, and our current clinical practices for voice assessment lack objective, quantifiable vocal effort measures in a manner that is both reliable and transferrable.[8]

Vocal effort means different things to different people; one patient's experience of effort may be internally calibrated differently than another patient, making these two experiences impossible to measure let alone compare.[7,9] Without the ability to systematically measure and compare ratings of vocal effort across patients, this frequently reported patient complaint can be left unmeasured and under-researched.

The mechanisms that contribute to vocal effort are as varied as they are complex, falling into one of the two categories: cognitive-behavioral influences and physiological influences. Cognitive effort and attention to the phenomenon of voicing is difficult to parse out from vocal effort, as outlined in a study by Vinney,[10] where vocal effort and cognitive effort covaried during low and high self-regulatory reading and writing tasks. Linguistic demands (eg, typological complexity or language dominance), motivation to communicate, and the social-emotional context of communication all contribute to perceived effort.[11] In fact, self-ratings of vocal effort are higher in negative mood states than positive mood states in normal speakers; this effect is greater in speakers with diagnosed voice disorders (VDs).[12]

Physiological contributions to vocal effort include physical activities used in speech production. Respiratory fitness and cardiovascular health impact the energy source of voice production, whereas articulatory precision and resonatory robustness

Accepted for publication August 28, 2014.

From the *Allied Health and Communicative Disorders, Northern Illinois University, DeKalb, Illinois; and the †Voice and Swallowing Clinic, University of Wisconsin, Madison, Wisconsin.

[1]Elizabeth Ford Baldner is now at University of Wisconsin Hospitals and Clinics, Madison, WI.

Address correspondence and reprint requests to Miriam van Mersbergen, Northern Illinois University, 1425 Lincoln Avenue, DeKalb, IL 60115. E-mail: mvanmersbergen@niu.edu

Journal of Voice, Vol. ■, No. ■, pp. 1-12
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2014.08.017

influence the filtered sound.[11] Phonatory and vibratory efficiency are determined by soft tissue dynamics including vocal fold stiffness,[5] presence of lesions,[5] internal temperature,[13] and hydration levels,[14–16] which all possess a documented connection with vocal effort. Muscular involvement may also influence perceived effort, as demonstrated by Dietrich and Verdolini Abbott,[17] where physical electromyography measures and vocal effort ratings both increased during periods of stress. Air volume, pressure, and flow parameters in conditions of fatigue and reduced hydration can also influence the perception of vocal effort.[14,15] However, physiological influences can be less direct, and during a vocally fatiguing task, phonation threshold pressure (PTP; a physiological measure) and perceived phonatory effort (PPE) presented with a moderately strong relationship. However, where PTP levels recovered almost immediately, PPE remained elevated for hours.[3] These findings have implications for the importance of rating-perceived vocal effort separately from PTP or other physiological measures.

Fatigue is another physical mechanism that can often be confounded with effort but nonetheless contributes to perceived exertion. The phenomenon of fatigue is defined as the degradation of peak capacity, or load on the physical system, over time.[18] Contrary to the purely physiological properties of fatigue, effort is a perceptual phenomenon that can be applied to various bodily functions, activities, and movements.[19] In other words, effort does not equal load; however, reduced load can have varying influences on perceived exertion.[20] Thus, the many possible cognitive-behavioral or physical phenomena connected to the perception of vocal effort creates a quagmire for the voice clinician and research concerned with vocal effort.

## Vocal effort scales used in the literature

Teasing out the mechanisms contributing to vocal effort is an ongoing challenge and identifying the most appropriate scale with which to measure vocal effort remains unclear. The varied investigations into vocal effort lack consistency in the quantification tools for vocal effort. Despite this, vocal effort is in need of a standard scale for the assessment of vocal effort.[5,8,21,22] The plethora of perceptual scales arising in the voice literature belies any consensus on how to capture perceived vocal effort. Measures reported in the literature thus far include direct magnitude estimation (DME) scales, ratio scales, visual analog scales (VASs), Likert-type scales, and category ratio scales. Each of these measures presents with unique outcomes in the context of measuring vocal effort.

The DME scales and interval scales have been used for subjective, psychophysical scaling in speech, and hearing science.[3,5,17,23–26] These scales require observers to assign numbers in proportion to the degree of value (or degree of effort experienced) and to correlate physical stimuli with corresponding numbers on a continuous scale. For example, consider a merit value of 10 on a DME scale; this dictates that twice as much merit equals a value of 20, and half as much merit equals a value of 5. What is most important in DME and interval scales is a direct connection to a physical, observable phenomena. In a study by Verdolini et al,[5] the researchers investigated the relationship between hydration level and PPE as measured with a DME scale. They found that perceived effort captured with a DME scale was less sensitive to changes in hydration level than PTP (a physical measure), suggesting that PPE as measured by DME is more complicated and one single measure cannot appropriately reflect vocal effort. In addition, the authors did not address inter- and intra-individual comparisons within their rating, highlighting one difficulty with DME scales: the requirement of a specific value or number in physical reality where individual's ratings are compared. Another difficulty of DMEs used to quantify vocal effort is the lack of comparison ability between subjects[27] because without a direct physical value or algorithm of numerous physical values, effort measures remain unstable between and within individuals.

Ratio scale intervals use equidistant points combined with an absolute zero point to measure a construct.[7] For example, a 40-year-old person is twice as old as a 20-year-old person, and zero means the absence of birth. The Voice Handicap Index-12 (VHI-12),[4,28] a condensed self-assessment of voice-related quality of life (including an item specifically addressing vocal effort) where the patient rates items according to a ratio scale (0-never, 1-hardly ever, 2-sometimes, 3-almost always, and 4-always), assumed that there is the same space between zero and one and there is between two and three. This form of ratio scaling is used in other studies of vocal effort in singers to categorize specific vocal tasks as high, medium, or low vocal effort.[29] However, although ratio scales provide precise, quantifiable data, there appears to be no known entity with a zero point. Neither enough is known about the construct of effort to determine the absolute zero point of effort nor is it possible to define effort with equal distance between all scale values as can be seen with physiological metric scaling.[7] Therefore, ratio scales fall short when considering vocal effort because effort appears to be a perceived entity as opposed to absolute entity that can be equally partitioned.

The VASs are psychometric scales in which subjective responses can be measured via positions along a continuous line with two defined endpoints. Unlike discrete scales such as ratio scales, and Likert-type scales, VASs allow for greater subjectivity and personalization of response because they allow an individual to respond without being constrained by imposed, equally divided responses.[30] The difficulty of these scales lies in the absence of actual numbers or values represented in physical reality; thus, comparison between subjects based on any external objectivity is impossible,[7] rendering clinical vocal effort measures ineffective. Nonetheless, VASs have been used by researchers to measure vocal effort in an effort to capture this construct with a physical measure.[1,13–15,22] However, without fixed units between each point in a given line, the results can neither be compared among individuals nor is there an objective, physical component to the measure, thus rendering interpretation of the data difficult.

Likert-type scales, another frequently used vocal effort measure, feature four or more points with fixed-choice responses and typically contain a neutral midpoint.[31–33] Like VASs, Likert-type scales do not require a concrete number in physical reality and therefore cannot be compared between subjects, thus rendering this scale unusable for research or clinical comparisons. Additionally, Likert-type scales give the illusion of distance equality between points on the scale, thereby taking perceptual phenomena (which are frequently nonlinear) and making them appear linear.[31] One superficial advantage to using Likert-type scales is that a researcher can use more than one characteristic to influence the perception of effort (eg, painful/tired or relaxed/clear). However, differences among multiple factors within individuals would skew clinical reporting using this scale (maybe a patient feels tired but not pain). This discrepancy further supports the need for a scale that incorporates both subjective and objective components[5] while simultaneously being anchored to well-defined endpoints with specific values for the intervals between rating values.

The description and quantification of effort is simultaneously subjective and objective.[5] It requires a psychometric scale that is both sensitive to perception and related to a physical measure. For this reason, scales that are either purely psychological (DME, Likert-type, and VAS scales) or purely physiological (ratio scales) are not complete in quantifying effort. A psychophysical scale, such as the category ratio scale,[7] combines the psychological and physical elements of effort into one measure, thus emerging as the most appropriate scale to research for vocal effort ratings. If clinicians continue to report on vocal effort with VD patients and researchers strive to converge on a consistent measurement of vocal effort, a scale must be validated for clinical use that covers both the subjective and objective factor of this construct. This scale should also boast interindividual reliability and validity for comparison among patients and between patient visits for progress monitoring, and should also be easy to administer and communicate to the lay population.

## A few scales by Gunnar Borg

As of yet, there is no tested scale that psychophysically measures the construct of vocal effort as perceived by the individual. In the realm of ergonomic science research, validated scales link the psychological and physiological responses of the human body to physical stress and exercise. One such scale is the Rating of Perceived Exertion (RPE) by Gunnar Borg.[33] The RPE is an exertion scale with ratings from six to 20; perceptual judgments listed on the scale range from “very, very light” to “very, very hard” and alternate every other number from seven to 19. This scale is a linear function of exercise and physical workload and is a good indicator of both physical stress and the psychological response to stress. In a study by Garcin and Billat,[34] the authors examined the relationships between perceived exertion and exercise duration during all-out runs. Two scales were compared, the RPE and the Estimation of Time Limit (ETL),[35] and showed that exercise duration (ETL) can be prescribed as a function of perceived exertion (RPE) and that the linear, psychophysical RPE scale was a good indicator of physical exertion. Making the link from physical exertion in exercise to physical exertion in voicing has not been studied, but for the purposes of seeking a validated scale for vocal effort, the RPE captures the necessary requirements for this construct of interest.

A study by Smit et al[36] highlighted a different psychophysical scale authored by Gunnar Borg. This study was concerned with vigilance and investigated whether two categories of effort—physical and mental—affected participants’ electroencephalograms and subjective alertness differently. The Borg scale,[5] constructed as both a category and ratio scale, was used as an indication of perceived physical load. The scale ranges from zero (not enduring) to 10 (very, very enduring), and the verbal expressions are placed on the scale according to their ratio properties (eg, if 4 represents “light,” then 2 should represent half of that value, both numerically and perceptively via the linguistic label). Although not directly related to the subject of vocal effort, this study showed how a Borg psychophysical scale was used to link together and quantify the physiological aspect of physical effort with the psychological aspect of mental effort. A link to the congruent aspects of speech and vocal effort is easily seen.

## The Borg Category Ratio-10 scale and vocal effort

Borg reiterates the importance of using standardized tools for evaluating subjective psychophysical symptoms and published yet another psychophysical scale[9] with applications in the perception of exertion and physical work. This new scale, the Category Ratio Scale (CR-10), is a method to rate perceived exertion. On the CR-10 scale, verbal expressions such as “nothing at all” to “extremely strong” are anchored to the correct positions on a logarithmic ratio scale from zero to 10, according to their quantitative meaning. The CR-10 is a “semi-public” unit (in other words, applicable to research, clinical, and lay settings) for interindividual comparisons among different groups of people and for different measured constructs. Because vocal effort is largely an individually subjective and perceptual phenomenon, the CR-10 is a useful psychometric scale for this purpose. It combines the logarithmic anchoring properties of many physical measures when compared with perceptual measures (as observed in many DME scales) and the individual perceptions captured in ratio scales and allows these units to be directly compared between individuals.

Other researchers have proposed the Borg CR-10 as a potential tool for measuring vocal effort and a few have suggested the Borg rating scale for future directions with their research.[3,11,37] The Borg CR-10 has been successfully used in clinical research for vocal effort ratings,[10,11,37,38] although it was not yet validated for such a purpose. It might be the most appropriate, empirically tested physiological and psychometric scale to use for vocal effort ratings, and it would fill the need for a diagnostic measure of vocal effort, both for research and clinical use. The question remains whether this scale can be validated for clinical use to measure vocal effort among voice patients, thus directing the present study.

## Vocal effort tasks

It is worth mentioning that within the literature reviewed for this study, measures of perceived exertion were always on a task immediately preceding the rating.[39–41] Vocal effort may be explored via a retrospective measure, such as with the VHI,[41] for the purposes of patient history gathering, and for qualitative information. In a study by Sampaio et al,[39] vocal effort was quantified retrospectively by a large cohort of teachers. They combined the vocal effort measure with other voice handicap measures to develop a long-term measure of vocal load among teachers. However, when reporting vocal effort measures for objective clinical and research purposes, it is vital that the measure be directly connected to a physical act (eg, a vocal loading task, acoustic protocol, or aerodynamic measure) and not a retrospective impression (eg, reporting on effort over the last few days, weeks, and year) to maintain the integrity of the construct of vocal effort.[14] Without anchoring the effort rating to a specific vocal task within a clinical or research protocol, it is difficult to accurately conclude that the perceived effort was specific to the patient's voice, not attributed to an overall sense of bodily or cognitive effort, or degraded by memory and emotional state. Protocols for vocal effort ratings, whereby raters judge a specific task in the moment, allow for more intimate and direct evaluation of effort rather than relying on one's memory for perceived effort in remote voicing scenarios. Additionally, in the context of clinical use of a vocal effort scale, linking effort ratings to already existing vocal tasks in an evaluation protocol would seem ecologically valid and time efficient.

### Purpose of the study

The purpose of this study is to determine whether or not the Borg CR-10 can measure vocal effort in patients with VDs. There are three main research questions, namely 1) Does the Borg CR-10 separate those with VDs from those without VDs?, 2) Does the Borg CR-10 vary with the severity of VD?, and 3) What vocal tasks best capture vocal effort using the Borg CR-10?

A clinically useful scale for vocal effort should correlate high vocal effort ratings with the presence of a VD. Voice clinicians must be able to use this scale together with a battery of other voice measures to accurately identify individuals with VDs. The scale for vocal effort must be both sensitive and specific to this patient population; thus, the first research question to answer is whether or not the Borg CR-10 separates those with VDs from those without VDs.

Although a binary measure of effort would be a useful diagnostic tool within the population of individuals with VDs, varying degrees of disorder severity exist. A scale that would capture the severity of perceived effort would add to the power of such a measure by giving patients more accurate readings and allowing clinicians to design evaluations and interventions tailored to severity levels. A vocal effort scale connected to the severity of VD is a requirement for validation; thus, the second research question is whether or not the Borg CR-10 varies with the severity of VD.

Clinical measurements of voice may be captured via a variety of vocal tasks, for example, sustained vowels, phonetically balanced or unbalanced read sentences or paragraphs, and conversational speech. The selection of the most appropriate tasks to capture vocal effort is paramount in the quest to define vocal effort with a clinical scale. It is unreasonable in a clinical setting to make effort ratings on every utterance; time constraints, limited resources, and patient attention and concentration to task are common clinical barriers to such a practice. Additionally, it is vital to get a live, real-time rating of an utterance and not simply a retrospective impression. Some utterances include other factors that may impact effort ratings; for example, effort might increase for articulation-laden sentences and conversation and decrease with vowels devoid of articulation. Clinicians and researchers require an accurate and efficient measure of vocal effort with the application of the fewest possible tasks; thus, the third research question is to determine what vocal tasks best capture effort using the Borg CR-10.

## METHODS

### Participants

Participants were members of two groups, namely a VD group and a healthy control (HC) group. All participants independently provided informed consent per Internal Review Board protocols at all participating institutions, were given the same training, and underwent the same protocol. No participants from either subject group received compensation for their participation.

**VD participants.** A total of 27 individuals with VDs, 10 males, 16 females, and one unidentified gender with average age of 50 years were recruited from the University of Wisconsin at Madison Voice and Swallow Clinics (UW-Madison) and the Northern Illinois University Speech-Language-Hearing Clinic (NIU). Inclusion criteria included adult patients with diagnoses of VDs including spasmodic dysphonia, functional dysphonia, benign mucosal lesions, peripheral neuropathies, and other organic and inorganic diseases of the vocal folds as documented by an otolaryngologist and a speech-language pathologist specializing in voice. Exclusion criteria included lack of a documented VD, aged younger than 18 years or older than 80 years, and non-English speaking.

**HC participants.** A total of 27 HCs, two males and 25 females, with an average age of 27 years, were recruited from the Northern Illinois University Speech-Language-Hearing Department. Exclusion criteria for the HC group included the presence or history of a VD, history of neurological impairment affecting speech or language, aged younger than 18 years or older than 80 years, and non-English speaking.

### Measures

**Stimuli.** A modified Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)[40] served as an independent measure. The modified CAPE-V tasks included sustained /a/, /i/, and /u/ vowels on a comfortable pitch and loudness; five standard read sentences, including 1) The blue spot is on the key again, 2) How hard did he hit him?, 3) We were away a year ago, 4) We eat eggs every Easter, 5) My mama makes lemon jam; and two questions or prompts that elicited 30 seconds of talking, which included the following: Tell me about your voice problem, and If you were to go on a vacation, where would you go and why?

**Validation measures.** *Voice Handicap Index.* The 30-item VHI[41] was used for construct validity and to verify participant group membership.

*Auditory-perceptual ratings.* Auditory-perceptual ratings were also used for construct validity and to verify participant group membership. Three expert judges, with an average of 20 years of experience in the field of voice and VDs rated 540 randomized utterances based on a the tasks on the CAPE-V.[40] Breathiness, roughness, and strain were rated on a 10-cm VAS via a computer program (Thomas Gleason, 2014) where judges mouse-clicked on the desired rating level from a 10-cm line presented via laptop screen. Room acoustics were measured from the location of each expert judge with a pure tone of 440 Hz, at 58-dB sound level. To assist with ease of listening, volume adjustments were made at the request of the expert judges if necessary. Training utterances included anchors for each of four quality characteristics (normal, breathy, rough, and strained) and four severity levels (normal, mild, moderate, and severe), for a total of 16 calibration ratings. After each training utterance, the judges discussed the utterance and arrived at a consensus on voice characteristics and severity level. They were retrained and recalibrated on these same utterances halfway through the session.

Inter-rater reliability for breathiness, roughness, and strain responses was calculated by correlating individual responses to the total average response. Correlations for vowels from each rater compared to the total of all vowel ratings ranged from $r(57) = 0.874$–$0.934$, $P < 0.001$ (two tailed). Correlations for sentences from each rater compared to the total of all sentence ratings ranged from $r(57) = 0.968$–$0.980$, $P < 0.001$ (two tailed). Correlations for conversation from each rater compared with the total of all conversation ratings ranged from $r(57) = 0.889$–$0.923$, $P < 0.001$ (two tailed). Inter-rater reliability for all utterance types (vowels, sentences, and conversation) indicated a strong reliability between the expert judges.

Intrarater reliability on 52 (10%) randomly chosen utterances ranged from $r(57) = 0.731$–$0.944$, $P < .001$ (two tailed) for breathiness, roughness, and strain, indicating a moderate-to-strong reliability for each expert judge.

*Phonation threshold pressure.* The PTP was included for criterion validity based on previous studies.[14,15]

**Measure of interest.** *Borg CR-10.* The adapted Borg CR-10[7] included scale ranges from zero (no vocal effort at all), 0.5 (very very slight vocal effort—just noticeable), 1 (very slight vocal effort), 2 (slight vocal effort), 3 (moderate vocal effort), 4 (somewhat severe vocal effort), 5 (severe vocal effort), 6–7 (very severe vocal effort), 8–9 (very very severe vocal effort—almost maximum), to 10 (maximum vocal effort; Appendix A). Participants selected the number on this scale that best represented their sense of effort when completing the CAPE-V and PTP tasks. Participants underwent the same training protocol for the adapted Borg CR-10.

## Instrumentation

The VHI, Borg CR-10, and SAM ratings were recorded with paper and pencil at both sites. For the participants run at UW-Madison, acoustic data were acquired via headset microphone (Micro-mic, C420, AKG, Northridge, CA) with a microphone-to-mouth distance of 10 cm and a microphone angle of 45° from corner to corner of the mouth. The data were recorded onto a hard drive via *Computerized Speech Laboratory* (*CSL*—KayPENTAX, Montvale, NJ) 4500. Aerodynamic (PTP) data were collected and analyzed via Phonatory Airflow System (PAS—KayPENTAX). For the participants run at NIU, acoustic data were acquired via headset microphone (Micro-mic, C520, AKG) with a microphone-to-mouth distance of 10 cm and a microphone angle of 45° from corner to corner of the mouth. The data were recorded onto a hard drive via *Computerized Speech Laboratory* 4500. Aerodynamic (PTP) data were collected and analyzed via Glottal Enterprises Aeroview Phonatory Aerodynamic System components (Glottal Enterprises, Syracuse, NY). The data were acquired via pneumotachometer airflow mask fitted with the OPA oral pressure adaptor and pressure and airflow transducers, digitized to an MS-110 analog digital converter, and sent to a Dell (Latitude, D83, Dell, Round Rock, TX) computer. The PTP data were captured and analyzed through *Aeroview* software.

## Procedure

After consenting to the study, participants were placed in an evaluation suite and filled out a general questionnaire about vocal hygiene, voice use, and presence of VD, as well as the paper-based VHI. Following training on how to interpret the Borg CR-10, participants produced the modified CAPE-V utterances. Immediately following each utterance, they rated their vocal effort using the Borg CR-10. Participants were then trained in the PTP protocol. Participants were initially trained to produce the softest possible phonation on the syllable /pi/ in modal pitch using multiple techniques to ensure the softest phonation possible. They were also trained in holding the mask to their face to ensure the tightest seal. Finally, they combined soft phonation with the mask. They produced five /pi/ syllables approximately 1 second apart. Following the PTP task, participants rated vocal effort on paper using the Borg CR-10. All tasks were counterbalanced to control for learning, emotion, and order effects.

## Statistical analysis

To confirm group separation between those with VDs and those without, independent-samples *t* tests were performed between the two groups on CAPE-V auditory-perceptual ratings (breathiness, roughness, and strain) for all utterances, the VHI scores, and PTP values. To identify whether the Borg CR-10 separates those with VDs from those without, independent-samples *t* tests were performed on average Borg CR-10 ratings for all utterances. To determine whether the Borg CR-10 is sensitive to the severity of VD, as well as to determine which tasks best capture vocal effort using the Borg CR-10, correlations between VHI, PTP, CAPE-V auditory-perceptual ratings (breathiness, roughness, and strain), and Borg CR-10 ratings were run for all utterances. All data were analyzed using *SPSS* statistical software (SPSS version 21, IBM, Armonk, NY).

## RESULTS

An independent-samples *t* test was performed to assess whether validation measures (mean auditory-perceptual ratings for breathiness, roughness, and strain, VHI, and PTP) differed significantly for VD participants compared with HCs. For all *t* tests results reported, all *P* values represent two-tailed tests and the pooled variance approach was used based on the equal variances assumption. For each validation measure (auditory-perceptual ratings, VHI, and PTP), independent-samples *t* tests were performed on all utterances: vowels, sentences, conversations, and the average of all utterances. Because there were no meaningful trends observed in any specific utterance and indeed all utterances produced similar results, only data on *t* tests from the average of each utterance will be presented.

### Auditory-perceptual ratings

**Breathiness.** Mean breathiness ratings for VD participants (M = 23.99, SD = 16.62) were about 17.87 points higher than the HC participants (M = 6.12, SD = 3.56), representing a significant difference, $t(52) = -5.46$, $P < 0.00$. The 95% confidence interval was −24.44 and −11.31 with a large effect size (Cohen's $d = 1.93$). These findings suggest that auditory-perceptual breathiness ratings significantly separate individuals with VDs from those without.

**Roughness.** Mean roughness ratings for VD participants (M = 25.57, SD = 18.51) were about 18.93 points higher than the HC participants (M = 6.64, SD = 2.50), representing a significant difference, $t(52) = -5.27$, $P < 0.00$. The 95% confidence interval was −26.14 and −11.72 with a large effect size (Cohen's $d = 1.97$). These findings suggest that auditory-perceptual roughness ratings significantly separate individuals with VDs from those without.

**Strain.** Mean strain ratings for VD participants (M = 25.30, SD = 22.58) were about 20.33 points higher than the HC participants (M = 4.97, SD = 2.22), representing a significant difference, $t(52) = -4.66$, $P < 0.00$. The 95% confidence interval was −29.10 and −11.57 with a large effect size (Cohen's $d = 1.79$). These findings suggest that auditory-perceptual strain ratings significantly separate individuals with VDs from those without. Please refer to Figure 1 for a graphical display of auditory perceptual ratings.

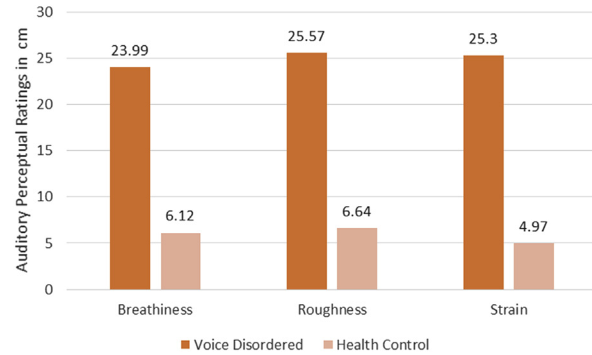


**FIGURE 1.** Auditory perceptual ratings for breathiness, roughness, and strain for participants with voice disorders and those without.

### Voice Handicap Index

An independent-samples *t* test was performed on mean VHI ratings for functional, physical, emotional, and total subscales. Mean VHI ratings on the functional subscale for VD participants (M = 12.14, SD = 10.00) were about 9.28 points higher than the HC participants (M = 2.86, SD = 3.18), representing a significant difference, $t(55) = -4.69$, $P < 0.00$. The 95% confidence interval was −13.25 and −5.31 with a large effect size (Cohen's $d = 1.42$).

Mean VHI ratings for the physical subscale for VD participants (M = 14.97, SD = 9.49) were about 12.22 points higher than the HC participants (M = 2.75, SD = 2.73), representing a significant difference, $t(55) = -6.55$, $P < 0.00$. The 95% confidence interval was −15.95 and −8.48 with a large effect size (Cohen's $d = 2.02$).

Mean VHI ratings for the emotional subscale for VD participants (M = 8.62, SD = 7.05) were about 7.44 points higher than the HC participants (M = 1.18, SD = 1.93), representing a significant difference, $t(55) = -5.39$, $P < 0.00$. The 95% confidence interval was −10.21 and −4.68 with a large effect size (Cohen's $d = 1.67$).

Mean total VHI ratings for VD participants (M = 35.72, SD = 24.92) were about 28.93 points higher than the HC participants (M = 6.79, SD = 6.32), representing a significant difference, $t(55) = -5.96$, $P < 0.00$. The 95% confidence interval was −38.67 and −19.21 with a large effect size (Cohen's $d = 1.87$). These findings suggest that the VHI significantly separates individuals with VDs from those without. Please refer to Figure 2 for a graphical display of VHI scores.

### Phonation threshold pressure

A final independent-samples *t* test was performed on mean PTP values. Mean PTP values for VD participants (M = 6.49, SD = 2.43) were about 3.37 points higher than the HC participants (M = 3.12, SD = 0.55), representing a significant difference, $t(49) = -7.15$, $P < 0.00$). The 95% confidence interval was −4.32 and −2.42 with a large effect size (Cohen's $d = 2.75$). Please refer to Figure 3 for a graphical display of PTP values.

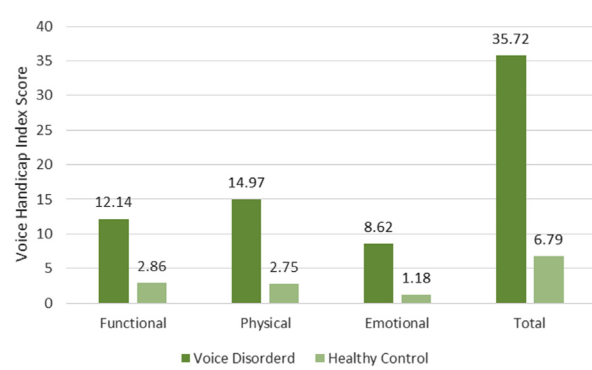


**FIGURE 2.** Voice Handicap Index scores for participants with voice disorders and those without.

### Borg CR-10 ratings

To answer the question if the Borg CR-10 is a sensitive measure to separate those with VDs from those without (HC), an independent-samples $t$ test was performed on the mean Borg CR-10 ratings for the vowels, sentences, conversation, the average of all utterances, and PTP (because Borg ratings followed PTP elicitation). All $t$ test results reported contained $P$-values representing two-tailed tests and used the pooled variance approach based on the equal variances assumption.

Mean Borg CR-10 vowel ratings for VD participants (M = 1.78, SD = 1.79) were about 0.40 points higher than the HC participants (M = 1.38, SD = 1.09) but failed to show significance, $t(56) = -1.01$, $P = 0.32$. The 95% confidence interval was −1.18 and 0.39 with a small effect size (Cohen's $d = 0.27$).

Mean Borg CR-10 sentence ratings for VD participants (M = 1.35, SD = 1.50) were about 0.09 points higher than the HC participants (M = 1.28, SD = 0.82) but failed to show significance, $t(56) = -0.22$, $P = 0.83$. The 95% confidence interval was −0.71 and 0.57 with a small effect size (Cohen's $d = .06$).

Mean Borg CR-10 conversation ratings for VD participants (M = 1.48, SD = 1.81) were about 0.03 points higher than the HC participants (M = 1.45, SD = 0.95) but failed to show significance, $t(56) = -0.08$, $P = 0.94$. The 95% confidence interval was −0.80 and 0.74. There was a small effect size (Cohen's $d = 0.02$).

Mean Borg CR-10 average of all utterances ratings for VD participants (M = 1.54, SD = 1.62) were about 0.17 points higher than the HC participants (M = 1.37, SD = 0.84) but failed to show significance, $t(56) = -0.49$, $P = 0.63$. The 95% confidence interval was −0.85 and 0.52 with a small effect size (Cohen's $d = 0.13$).

Mean Borg CR-10 PTP ratings for VD participants (M = 2.10, SD = 1.89) were about 0.17 points higher than the HC participants (M = 1.93, SD = 1.28) but failed to show significance, $t(56) = -0.40$, $P = 0.69$. The 95% confidence interval were −1.03 and 0.68 with a small effect size (Cohen's $d = .11$). These findings suggest that the Borg CR-10 fails to show significant separation between individuals with VDs and those without. Please refer to Figure 4 for a graphical display of Borg CR-10 scores.

### Correlations

To answer the question if the Borg CR-10 co-varies with severity of VD, Pearson bivariate correlations for each utterance type (vowels, sentences, conversations, and the average of all utterances) were performed on each validation measure (auditory perceptual ratings, VHI, and PTP). Because there were no meaningful trends observed in any specific utterance and indeed all utterances had produced similar results, data on Pearson bivariate correlations from the average of all utterances will be presented. For the Borg CR-10 ratings, VHI, and PTP, correlations were performed on the total sample of 54 participants: VD and HC combined.

Pearson correlations were calculated among the Borg CR-10, auditory-perceptual ratings (breathiness, roughness, and strain), and VHI ratings (Table 1). All correlations among the Borg CR-10, auditory-perceptual ratings, and the VHI were statistically significant and correlations ranged from mild to high correlations. Please refer to Table 1 for correlations and significance levels among Borg CR-10, auditory perceptual ratings, and VHI scores.

Borg CR-10 ratings for PTP and actual PTP values were correlated. The correlation between Borg CR-10 ratings for PTP and actual PTP values was not statistically significant, $r(51) = 0.03$, $P < 0.858$ (two tailed). Please refer to Table 2 for correlation of PTP values with the corresponding Borg CR-10 rating.

Finally, to answer the question if vocal effort levels reported in the Borg CR-10 could be best predicted by specific types of utterances, Pearson correlations were calculated among the Borg CR-10 ratings for vowels, sentences, conversation, and average of all utterances. The correlations among Borg CR-10 ratings for vowels, sentences, conversation, and average of all utterances were all statistically significant with high correlations. Please refer to Table 3 for correlations and significance levels of Borg CR-10 values for vowels, sentences, conversation, and the average of all utterances.

## DISCUSSION

This preliminary study sought to determine whether or not the Borg CR-10 is a useful tool to measure vocal effort in patients

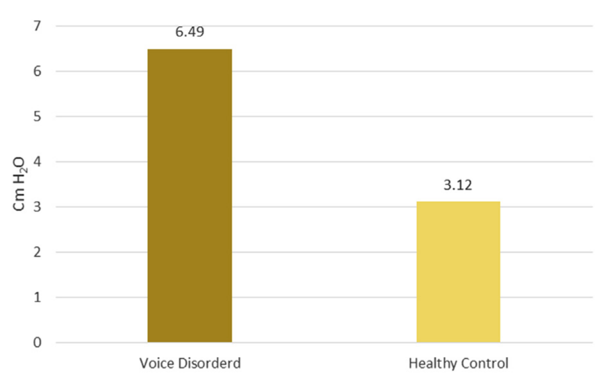


**FIGURE 3.** Phonation threshold pressure values for participants with voice disorders and those without.

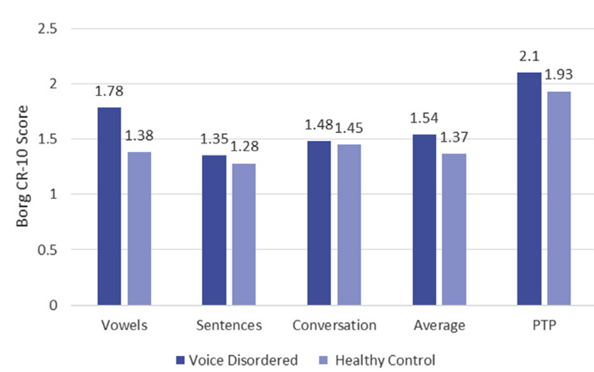


**FIGURE 4.** Borg CR-10 ratings for vowel, sentence, conversation, average of all utterances, and phonation threshold pressure for participants with voice disorders and those without.

**TABLE 1.**
**Correlations Among Borg CR-10, Auditory Perceptual Ratings of Breathings, Roughness, and Strain, and the Voice Handicap Index**

| Parameters | Borg CR-10 | Breathiness | Roughness | Strain | VHI |
|---|---|---|---|---|---|
| Borg CR-10 | | | | | |
| Pearson correlation | 1 | 0.358* | 0.381* | 0.455* | 0.539* |
| Significance (two tailed) | | 0.008 | 0.004 | 0.001 | 0.000 |
| N | 54 | 54 | 54 | 54 | 54 |
| Breathiness | | | | | |
| Pearson correlation | 0.358* | 1 | 0.919* | 0.931* | 0.652* |
| Significance (two tailed) | 0.008 | | 0.000 | 0.000 | 0.000 |
| N | 54 | 54 | 54 | 54 | 54 |
| Roughness | | | | | |
| Pearson correlation | 0.381* | 0.919* | 1 | 0.955* | 0.570* |
| Significance (two tailed) | 0.004 | 0.000 | | 0.000 | 0.000 |
| N | 54 | 54 | 54 | 54 | 54 |
| Strain | | | | | |
| Pearson correlation | 0.455* | 0.931* | 0.955* | 1 | 0.621* |
| Significance (two tailed) | 0.001 | 0.000 | 0.000 | | 0.000 |
| N | 54 | 54 | 54 | 54 | 54 |
| VHI | | | | | |
| Pearson correlation | 0.539* | 0.652* | 0.570* | 0.621* | 1 |
| Significance (two tailed) | 0.000 | 0.000 | 0.000 | 0.000 | |
| N | 54 | 54 | 54 | 54 | 54 |

* Correlation is significant at the 0.01 level (two tailed).

with VDs. Three main research questions were answered in this study. The first question asked if the Borg CR-10 separates those with VDs from those without VDs. The results of *t* tests on all vocal tasks indicated that the Borg CR-10 does not significantly separate individuals with VDs from those without VDs. The second question examined if the Borg CR-10 varies with the severity of VD. The bivariate correlation of Borg CR-10, auditory-perceptual ratings, PTP, and VHI scores indicated that the Borg CR-10 moderately varies with the severity of VD. The final question, what vocal tasks best capture vocal effort using the Borg CR-10, was answered via bivariate correlations among the Borg CR-10 ratings of vowels, sentences, conversation, and average of all utterances. Results indicated that there were strong correlations among all vocal tasks used in this study, suggesting that no single vocal task captures vocal effort differently.

It appears as though the Borg CR-10 falls short in its usefulness for measuring vocal effort. Discrete numerical values attributed to the scale may present the largest limitation. Groups means of both VD and HC fell between one and two (1.3 compared with 1.7), suggesting that if the scale is to be completed as intended, this subtle difference would be unobserved. The discrete nature of the Borg CR-10 would make it impossible for an individual to choose, for example 1.3, therefore ratings could not accurately separate groups. Incorporating intermediate anchors that fall between main anchors or adding more specific definitions to anchors may possibly alleviate this issue, but further research on the validity or feasibility of adding sublevels is required.

Another limitation with this version of the Borg CR-10 scale is that anchors may not be in keeping with the range of effort used in voicing. Perhaps individuals are incapable of feeling "no vocal effort" or "maximum vocal effort," rendering these endpoints on the Borg CR-10 meaningless. The difficulty in assessing perceived exertion with voicing is that vocal effort sensations have yet to be well defined and possible anchors have yet to be empirically validated. This point cannot be stressed enough and calls into question whether or not a completely new scale category ratio should be designed for collecting vocal effort ratings.

The Borg CR-10 scale was designed to measure large groups of muscles for heavy physical exertion, such as running and weight-bearing activities. Because voicing uses far less than the maximum exertion, perhaps vocal effort is simply too subtle to be detected by this scale. For example, vocal effort may be more about coordination than about total force during voicing. Any measure that addresses vocal effort would need to consider the unique aspects of effort with respect to multiple aspects of voicing compared with physical exertion.

**TABLE 2.**
**Correlations Between Phonation Threshold Pressure and the Corresponding Borg CR-10 Rating**

| Parameters | Borg CR-10 | PTP |
|---|---|---|
| Borg CR-10 | | |
| Pearson correlation | 1 | 0.026 |
| Significance (two tailed) | | 0.858 |
| N | 54 | 54 |
| PTP | | |
| Pearson correlation | 0.026 | 1 |
| Significance (two tailed) | 0.858 | |
| N | 54 | 54 |

**TABLE 3.**
**Correlations Among Borg CR-10 Ratings for Vowel, Sentence, Conversation, and the Average of All Utterances**

| Parameters | Vowels | Sentences | Conversation | Average |
|---|---|---|---|---|
| Vowels | | | | |
| Pearson correlation | 1 | 0.871* | 0.753* | 0.936* |
| Significance (two tailed) | | 0.000 | 0.000 | 0.000 |
| N | 54 | 54 | 54 | 54 |
| Sentences | | | | |
| Pearson correlation | 0.871* | 1 | 0.836* | 0.957* |
| Significance (two tailed) | 0.000 | | 0.000 | 0.000 |
| N | 54 | 54 | 54 | 54 |
| Conversation | | | | |
| Pearson correlation | 0.753* | 0.836* | 1 | 0.921* |
| Significance (two tailed) | 0.000 | 0.000 | | 0.000 |
| N | 54 | 54 | 54 | 54 |
| Average | | | | |
| Pearson correlation | 0.936* | 0.957* | 0.921* | 1 |
| Significance (two tailed) | 0.000 | 0.000 | 0.000 | |
| N | 54 | 54 | 54 | 54 |

* Correlation is significant at the 0.01 level (two tailed).

Lastly, given the nature of personal perceptions, even well-validated anchors might require extensive defining to ensure consensus of meaning for each anchor. Additionally, external verification of understanding of these well-defined anchors should be used to determine if an individual is ready to rate his or her vocal effort. Anchors for this version of the Borg CR-10 were vague and individual variability in interpretation of each anchor might have been controlled with more clearly defined, externally verified anchors. Even small differences in training protocols might yield vastly different vocal effort rating results, rendering the data useless for between-subject comparisons. Administrator training would invariably lead to better results. Further studies on various training regimens should be explored for the most valid, efficient, and appropriate effort ratings possible.

### Clinical and research implications for capturing vocal effort

One remarkable outcome from this study is that the Borg CR-10 is not successful in separating out disordered and healthy voices despite moderate correlations with measures that do. On the surface, and based on this outcome alone, the Borg CR-10 should not be used for clinical or research purposes as a diagnostic tool. However, for research purposes, the Borg CR-10 may be functional if the intent is to measure severity of VD as a construct, as evidenced by the significant findings from research projects using the Borg CR-10 scale.[11,36,37] However, for clinical assessment where the intent is to identify the presence of a VD, this scale falls short.

Another question arising from these findings relates to individual differences in perceptions of vocal effort. Are those with VDs more aware of effort, given that this is a salient feature of their experience, or are they less aware, which predisposed them to the VDs in the first place? Conversely, are HCs more conscious of small changes in effort and therefore modify vocal behavior that protects vocal functioning, or are they less sensitive to effort because they do not need to be? Differing sensitivity levels of vocal effort between these groups may have been one reason this scale did not differentiate them. A vocal effort scale should either accommodate both groups of individuals or only be used within one group for research purposes.

Another point to mention is that the perception of vocal effort appears to be a trained phenomenon. According to research by van Leer,[38] voice patients rated vocal effort lower on initial evaluation measures before voice therapy. They presented with subsequent increases in vocal effort ratings once they began therapy. This finding is at odds with what we would expect given that if the voice therapy goals are to improve phonatory efficiency, effort would hopefully decrease, not increase. Perhaps, these patients learned awareness of their internal sense of effort and once they had a meaningful context by which to recognize effort, they were able to rate it more accurately. Thus, training for perception of effort may be necessary to obtain valid effort ratings.[23,32,37] Without training, patients do not have an anchor for their perception of minimal and maximal effort, and any administered self-perception scale becomes meaningless as a pre- and post-training measure.

A final, but important consideration on using the Borg CR-10 clinically is that the utterances in this research protocol may not have been stressful enough to require sufficient and noticeable effort from the vocal mechanism. We chose the investigated utterances to preserve clinic flow and avoid adding more utterances to a voice evaluation. To this end, we used typical utterances from the CAPE-V and PTP measures. Quite possibly, sustaining vowels for 5 seconds, reading five sentences, answering two simple questions, and phonating at threshold softness on /pi/ do not uncover changes or differences in vocal effort to a measurable degree. More vocally challenging tasks, such as reading multiple paragraphs, loud phonation, or singing at 80% of the vocal range, might be more appropriate to capture vocal effort.[14,15] Conceptualizing vocal effort in the context of a stress test similar to a cardiac stress test, where vocal effort is measured after sustained use at a percentage of maximum stress, might yield more meaningful group differences. Evidence that this might be necessary is observed in previous voice research that has successfully measured vocal effort where stressful and fatiguing utterances were part of the research protocols.[14,15,23] Future research on effort scale validation should consider utterance type as a factor determining group differences or establishing severity ratings via effort ratings. One possibility would be loud reading until a specific reported effort level was reached. The longer an individual reads would indicate a more robust system. This area of effort rating validation is rife with further study possibilities.

## Convergent validity of the Borg CR-10

Correlations between the Borg CR-10, VHI, and auditory-perceptual ratings across vocal tasks were all statistically significant and moderately high. Although the Borg CR-10 does not differentiate HC and VD groups, it may be useful alongside a battery of other diagnostic voice measures. The Borg CR-10 might add another dimension to voice assessment data that does not currently exist. However, although the Borg CR-10 may be useful for measuring severity based on VHI and auditory-perceptual correlations, there was no statistical significance to the correlation between the Borg CR-10 and PTP. This result is especially puzzling because previous studies have revealed a correlation between PTP and PPE.[3,5,14,15] All of these studies elicited PTP at the 80% range—which was different than this study that measured PTP at normal pitch and normal loudness, lending further support to investigate the vocal task with which to measure perceived vocal effort. However, there are caveats in task elicitation that would require careful consideration. The more complicated a vocal elicitation task, for example PTP at 80% of physiologic range, may confound vocal effort ratings with the difficulty required to complete the task. Vinney[10] observed this type of phenomenon and noted difficulty differentiating cognitive effort from vocal effort.

## Limitations to the research

Sample size was the biggest limitation to this study. The effect sizes were small for all the Borg CR-10 comparisons (Cohen's $d = 0.01–0.26$) based on the sample size of 54 participants. It was not appropriate or feasible for the scope of this study to run more participants to increase power. Taking into account these effect size limitations, the results of this study are preliminary and more participants are needed to properly analyze separation of groups.

Another limitation to this study was the omission of other validated voice measures in the research protocol. Acoustic voice measures that reflect phonatory efficiency, additional aerodynamic voice measures such as laryngeal airway resistance, and other physiologic vocal measures such as electroglottography, may have yielded additional quantifiable data for analysis against the Borg CR-10 and may have assisted with determining relative contributions to perceived vocal effort.[42]

A final limitation to this study was the poor balance of age and male-to-female ratio among participants between the voice-disordered group and the HC group.

## Future directions

In addition to determining the best task to elicit vocal effort ratings, as stated earlier, other considerations may help shape a valid measure of vocal effort. Because the Borg CR-10 was validated on large muscle groups and not on coordination and fine motor exertion for voicing, future investigations should find ways in which to include these constructs into perceived effort ratings. In addition, future studies should attempt to control ratings of perceived vocal effort with individual differences in internal awareness.

Future studies should look into whether the Borg CR-10 differentiates various VDs. Does the Borg CR-10 separate groups within the VD population? In other words, can effort ratings via the Borg CR-10 differentiate between specific VDs, such as spasmodic dysphonia and muscle tension dysphonia? In this study, all VD diagnoses were collapsed together and may have yielded variability too wide to elicit the most sensitive results. Future studies might want to address only one or two diagnostic groups to provide valuable differential diagnosis to the clinic in a similar way that auditory-perceptual ratings differ for adductor spasmodic dysphonia versus muscle tension dysphonia.[43] However, until proper tasks to elicit vocal effort are identified, this question remains elusive.

Finally, other formats of recording vocal effort ratings using interactive, digitized media should be explored. Such methods would undoubtedly lend efficiency to clinic flow and accuracy to research protocols.

## Acknowledgments

This research was part of the requirements for the master's degree in speech-language pathology for the first author and was conducted under the supervision of Miriam van Mersbergen at Northern Illinois University. The authors would like to thank Dr. Susan Thibeault and the speech-language pathologists at the University of Wisconsin at Madison Voice and Swallow Clinics who helped us collect and organize the participant data. They would also like to thank the expert judges for their time and keen ears—Michelle Eppley, Liz Platt, and Kate DeVore. Finally, they would like to thank Regina Dentzman, Rachael Clark McCarthy, and Asra Razvi for their research assistance and the anonymous reviewers for their helpful comments.

## REFERENCES

1. Stepp CE, Sawin DE, Eadie TL. The relationship between perception of vocal effort and relative fundamental frequency during voicing offset and onset. *J Speech Lang Hear Res*. 2012;55:1887–1896.
2. Lagier A, Vaugoyeau M, Ghio A, Legou T, Giovanni A, Assaiante C. Coordination between posture and phonation in vocal effort behavior. *Folia Phoniatr Logop*. 2010;62:195–202.

3. 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.
4. Hanschmann H, Lohmann A, Berger R. Comparison of subjective assessment of voice disorders and objective voice measurement. *Folia Phoniatr Logop*. 2011;63:83–87.
5. Verdolini K, Titze IR, Fennell A. Dependence of phonatory effort on hydration level. *J Speech Hear Res*. 1994;37:1001–1007.
6. Liénard J, Di Benedetto M. Effect of vocal effort on spectral properties of vowels. *J Acoust Soc Am*. 1999;106:411–422.
7. Borg G. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc*. 1982;14:377–381.
8. Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *J Voice*. 2004;18:299–304.
9. Borg G. Psychophysical scaling with applications in physical work and the perception of exertion. *Scand J Work Environ Health*. 1990;16:55–58.
10. Vinney L. *The role of self-regulation in the modification of vocal behavior during reading and speech* [dissertation]. Madison, WI: University of Wisconsin, Madison; 2013.
11. van Mersbergen M, Patrick C, Glaze L. Functional dysphonia during mental imagery: testing the train theory of voice disorders. *J Speech Lang Hear Res*. 2008;51:1405–1423.
12. Ferrand CT. *Speech Science: An Integrated Approach to Theory and Clinical Practice*. 2nd ed. Boston, MA: Allyn & Bacon; 2007.
13. Sandage MJ, Connor NP, Pascow DD. Voice function differences following resting breathing versus submaximal exercise. *J Voice*. 2013;27:572–578.
14. Solomon NP, Stemmle DiMattia M. Effects of a vocally fatiguing task and systemic hydration on phonation threshold pressure. *J Voice*. 2000;14:341–362.
15. Solomon NP, Glaze LE, Arnold RR, van Mersbergen M. Effects of a vocally fatiguing task and systemic hydration on men's voices. *J Voice*. 2003;17:31–46.
16. Jiang J, Ng J, Hanson D. The effects of rehydration on phonation in excised canine larynges. *J Voice*. 1999;13:51–59.
17. Dietrich M, Verdolini Abbott K. Vocal function in introverts and extraverts during a psychological stress reactivity protocol. *J Speech Lang Hear Res*. 2012;55:973–987.
18. Bogdanis GC. Effects of physical activity and inactivity on muscle fatigue. *Front Physiol*. 2012;3:142.
19. Hollander DB, Kilpatrick MW, Ramadan ZG, et al. Load rather than contraction type influences rate of perceived exertion and pain. *J Strength Cond Res*. 2008;22:1184–1193.
20. Cafarelli E, Cain WS, Stevens JC. Effort of dynamic exercise: influence of load, duration, and task. *Ergonomics*. 1977;20:147–158.
21. Eriksson A, Tranmüller H. Perception of vocal effort and distance from the speaker on the basis of vowel utterances. *Percept Psychophys*. 2002;64:131–139.
22. Laukkanen A, Ilomäki I, Leppänen K, Vilkman E. Acoustic measures and self-reports of vocal fatigue by female teachers. *J Voice*. 2008;22:283–289.
23. Hunter EJ, Titze IR. Quantifying vocal fatigue recovery: dynamic vocal recovery trajectories after a vocal loading exercise. *Ann Otol Rhinol Laryngol*. 2009;118:449–460.
24. Schiavetti N, Sacco PR, Metz DE. Direct magnitude estimation and interval scaling of stuttering severity. *J Speech Hear Res*. 1983;26:568–573.
25. Verdolini-Marston K, Burke MK, Lessac A, Glaze L, Caldwell E. Preliminary study of two methods of treatment for laryngeal nodules. *J Voice*. 1995;9:74–85.
26. Rosenthal AL, Lowell SY, Colton RH. Aerodynamic and acoustic features of vocal effort. *J Voice*. 2014;28:144–153.
27. Eadie TL, Doyle PC. Direct magnitude estimation and interval scaling of pleasantness and severity in dysphonic and normal speakers. *J Acoust Soc Am*. 2002;112:3014–3021.
28. Nawka T, Verdonck-de Leeuw IM, De Bodt M, et al. Item reduction of the voice handicap index based on the original version and on European translations. *Folia Phoniatr Logop*. 2009;61:37–48.
29. 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.
30. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as a ratios scale measures for chronic and experimental pain. *Pain*. 1983;17:45–56.
31. Warner RM. *Applied Statistics: From Bivariate Through Multivariate Techniques*. Thousand Oaks, CA: Sage Publications; 2008:255–278.
32. Thomas-Kersting C, Casteel RL. Harsh voice: vocal effort perceptual ratings and spectral noise levels of hearing-impaired children. *J Commun Dis*. 1989;22:125–135.
33. Borg G. Perceived exertion as an indicator of somatic stress. *Scand J Rehabil Med*. 1970;2:92–98.
34. Garcin M, Billat V. Perceived exertion scales attest to both intensity and exercise duration. *Percept Mot Skills*. 2001;93:661–671.
35. Garcin M, Vandewalle H, Monod H. A new rating scale of perceived exertion based on subjective estimation of exhaustion time: A preliminary study. *Int J Sports Med*. 1999;20:40–43.
36. Smit AS, Eling PA, Hopman MT, Coenen AML. Mental and physical efforts affect vigilance differently. *Int J Psychophysiol*. 2005;57:211–217.
37. Steinhauer K, Grayhack JP, Smiley-Oyen AL, Shaiman S, McNeil MR. The relationship among voice onset, voice quality, and fundamental frequency: a dynamical perspective. *J Voice*. 2004;18:432–442.
38. van Leer E. *Effect of social cognitive variables in voice therapy outcomes* [dissertation]. Madison, WI: University of Wisconsin, Madison; 2010.
39. Sampaio MC, Borges dos Reis EJF, Carvalho FM, Porto LA, Araujo TM. Vocal effort and voice handicap among teachers. *J Voice*. 2012;26:820.e15–820.e18 [Retrospective ratings].
40. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kramer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18:124–132.
41. Jacobson BH, Johnson A, Grywalski C, et al. The Voice Handicap Index (VHI): Development and validation. *Am J Speech Lang Pathol*. 1997;6:66–70.
42. Baken RJ, Orlikoff RE. *Clinical measurement of speech and voice*. San Diego, CA: Singular Publishing Group; 2000.
43. Roy N, Gouse M, Mauszycki SC, Merrill RM, Smith ME. Task specificity in adductor spasmodic dysphonia versus muscle tension dysphonia. *Laryngoscope*. 2005;115:311–316.

**APPENDIX A.**
**Adapted BORG CR-10 for Vocal Effort Ratings**

| Severity | Scale |
|---|---|
| No vocal effort at all | 0 |
| Very very slight vocal effort (just noticeable) | 0.5 |
| Very slight vocal effort | 1 |
| Slight vocal effort | 2 |
| Moderate vocal effort | 3 |
| Somewhat severe vocal effort | 4 |
| Severe vocal effort | 5 |
| | 6 |
| Very severe vocal effort | 7 |
| | 8 |
| Very very severe vocal effort (almost maximum) | 9 |
| Maximum vocal effort | 10 |