# Effects of Vocal Function Exercises: A Systematic Review

**Vrushali Angadi, Daniel Croake, and Joseph Stemple,** *Lexington, Kentucky*

**Summary: Objectives.** The purpose of the present review was to systematically analyze the evidence for the effectiveness of vocal function exercises (VFEs) in improving voice production.

**Methods.** A systematic literature search was performed by two independent reviewers using PubMed and EBSCOHost to access relevant databases and to locate outcome studies that used VFEs as an intervention. Articles that met inclusion criteria were appraised based on the American Speech-Language and Hearing Association's levels of evidence. Effect sizes for outcomes were calculated using Hedge's g. Voice outcomes were categorized according to the five domains of voice assessment: visual perceptual analysis, acoustic analysis, aerodynamic analysis, auditory-perceptual analysis, and patient self-report measures.

**Results.** Twenty-one articles were included for the final appraisal. All studies demonstrated positive effects of VFEs as demonstrated by effect sizes across selected voice parameters. Effect sizes across parameters ranged from –0.59 to 1.55. None of the included studies reported adverse voice outcomes as a result of VFEs.

**Conclusions.** Outcome studies demonstrate that VFEs are efficacious in enhancing vocal function in individuals with normal and disordered voices, presbylaryngeus, and professional voice users. The available research suggests moderate to strong evidence to support the use of VFEs for a variety of voice disorders.

**Key Words:** Vocal function exercises–Systematic review–Effect sizes–PICO–Voice therapy.

## INTRODUCTION

The number of peer-reviewed studies using vocal function exercises (VFEs) as a primary treatment for voice disorders has significantly increased over the past few years. VFEs were developed as a physiologic approach for treating voice disorders via direct manipulation of the underlying anatomy and physiology important for healthy voice production. Clinically, these exercises are thought to promote a relative physiologic balance among the subsystem triad of respiration, phonation, and resonance by strengthening the intrinsic laryngeal and auxiliary musculature, and also by enhancing vocal tract resonance characteristics.[1] To date, 27 peer-reviewed papers have demonstrated outcomes using the VFE program.[2–28] Data suggest that VFEs improve or remediate voice problems in both young and aging adults,[2,5,6,8,9,12,13,15–17,19,22–24,26–28] enhance normal voice production,[3,18] provide preventative protection to persons with high-dose vocal demands,[11,20] and enhance vocal function in graduate-level operatic singers.[7,21]

As with any medical treatment, it is important to critically appraise the available evidence to make informed decisions about patient care. Sackett et al[29] defined evidence-based practice as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients [by] integrating individual clinical expertise with the best available external clinical evidence from systematic research." Evidence for voice therapy in general has been systematically reviewed; however, it was concluded that direct comparisons across voice therapy types made attempts to interpret and generalize the information difficult. A solution to this problem would be to isolate and analyze the evidence for one particular voice therapy method.[30] Because a relatively large number of studies have used VFEs as a primary intervention, a comprehensive systematic review is warranted to determine the current state of evidence for this particular therapy program, and to provide information to clinicians on appropriate use of VFEs with their patients. The purpose of this paper was to determine the current state of evidence for the VFE program for the treatment of voice disorders and to determine their effectiveness as measured by one or more of the five domains of voice assessment encompassing audioperceptual, client perceived, acoustic, aerodynamic, and laryngeal imaging techniques.

### Epidemiology

The prevalence of voice disorders in the United States is 29%.[31] This number is markedly increased for professional voice users, reaching nearly 50%.[25,32] Voice disorders have been shown to affect social, psychological, emotional, physical, and functional aspects of communication and health[33] as well as job satisfaction, performance, and attendance.[33–35] A study published in 2001 stated that the societal cost of voice problems for teachers alone was roughly 2.5 billion dollars (3.48 billion when adjusted for inflation in 2016) in health care and absentee compensation annually.[36] Voice disorder claims account for about $3400 in disability payments and as many as 40 lost workdays per claim.[37] These losses are comparable with other chronic health conditions such as heart disease, asthma, and depression.[37] Voice impairments are critical functional limitations in the current communication age and are a significant public health concern.

### Physiologic voice therapy

Physiologic voice therapy is a holistic approach to remediate voice disorders by directly manipulating the underlying physical components of voice production.[38] This is in contrast to hygienic

Accepted for publication August 31, 2017.

From the Division of Communication Sciences and Disorders, University of Kentucky, 900, South Limestone, Suite 120, Lexington, Kentucky.

Address correspondence and reprint requests to Vrushali Angadi, Division of Communication Sciences and Disorders, University of Kentucky, 900, South Limestone, Suite 120, Lexington, KY 40536-0200. E-mail: vrushali.angadi@gmail.com

Journal of Voice, Vol. 33, No. 1, pp. 124.e13–124.e34

0892-1997



https://doi.org/10.1016/j.jvoice.2017.08.031

therapy that seeks to reduce phonotraumatic behaviors and increase systemic hydration, and symptomatic voice therapy that is based on the modification of observed vocal symptoms.

Physiologic voice therapy currently has the strongest evidence base. A review by Thomas and Stemple found that physiologic approaches to voice therapy were superior to both symptomatic and vocal hygiene therapies when used as standalone approaches.[39] VFEs in particular were developed to simultaneously target the disrupted physiologic components in the vocalization subsystems of respiration, phonation, and resonance to restore their physiologic balance.[18]

Determining the evidence for voice therapy in general has proven difficult because of the multidimensional nature of voice production, the broad scope of therapeutic approaches, differences in outcomes measures, differences in instrumentation, and due to heterogeneous populations and types of disorders.[30] To strengthen the ability to generalize results, we have limited the scope of this review to only VFEs because it is delivered in a systematic and universal manner.

## The five domains of voice assessment

Because voice production is the *product* of the interactions of respiration, phonation, and resonance, it is a multidimensional event and may be measured qualitatively and quantitatively by various means. The five common domains of voice assessment are patient perceived self-rating scales, auditory-perceptual scales, acoustic, aerodynamic, and visual perceptual assessments. Hirano[40] stated, "Voice is multidimensional in nature, so we need a set of tests to evaluate function in its entirety." This has been reiterated by Titze (Titze, 1991 in Bless and Hicks[41]) who stated, "Diagnostic hypotheses should not be made on basis of one test or measure because one cannot look at an isolated phenomenon without running the risk of misinterpreting the results".[41] Multidimensional assessment of voice helps overcome the limitations of any one assessment type.[41]

Meaningful voice assessments are predicated by their validity and reliability and should have precedence in the literature with a strong evidence base in and of themselves. To strengthen interpretation of our findings, we have calculated effect sizes for measures that were common across studies where available. A thorough review of these assessment measures is beyond the scope of this paper; however, it is prudent to provide a brief summary of the validity and reliability of some common measures used to strengthen the interpretation of our findings.

## Auditory perceptual measures

### *Consensus Auditory Perceptual Evaluation of Voice*

The Consensus Auditory Perceptual Evaluation of Voice (CAPE-V) was developed in 2002 at the American Speech-Language and Hearing Association consensus meeting.[42] This tool was developed to describe the severity of auditory-perceptual attributes of a voice problem in a way that can be communicated among clinicians, and to contribute to hypotheses regarding the anatomic and physiological bases of voice problems and to evaluate the need for additional testing. This evaluation has demonstrated criterion validity and both intra- and inter-rater reliability.[43,44]

### *GRBAS scale*

The GRBAS scale[45] was developed to rate vocal quality within five perceptual categories: overall grade (G), roughness (R), breathiness (B), asthenia (A), and strain (S). Validity and reliability data have been established in multiple studies (eg, references 44,46–49).

## Patient-perceived rating scales

### *Voice Handicap Index*

The Voice Handicap Index (VHI), developed by Jacobson et al,[50] is a 30-item questionnaire designed to assess the patient-perceived impact of a voice disorder in three domains: physical, emotional, and social. Validity and reliability of this tool are reported in the study of Jacobson et al.

### *Voice-Related Quality of Life*

The Voice-Related Quality of Life (V-RQOL)[51] is a 10-item questionnaire that also probes patient perceived difficulty with their voice in physical and socio-emotional domains. Validity and reliability of this instrument have been established (eg, references 51,52).

## Acoustic measures

Because there are many acoustic measures available, we will review three common measures used across studies included in this review. Each measure provides change scores demonstrating post-therapeutic voice changes.

### *Jitter and shimmer*

Jitter is the short-term cycle-to-cycle variation in frequency in a voice sample, whereas shimmer is the short-term cycle-to-cycle variation in amplitude. Both of these measures have failed to demonstrate strong test-retest reliability because of differences in extraction methods across systems and because highly dysphonic voice signals decrease the reliability of these measures.[53,54] Validity of these measures has been demonstrated by attempting to correlate these measures with perceptually rated measures of voice quality.[46,55]

### *Harmonic-to-noise ratio*

Harmonic-to-noise ratio (HNR) is based on the premise that vocal production consists of a strong harmonic component with a smaller degree of aperiodic noise.[56] Voices that carry a stronger harmonic component compared with the noise component should yield better voice quality. Validity and reliability data for this measure have been determined and have been demonstrated to be more robust than jitter and shimmer (eg, references 57–59).

## Aerodynamic measures

Aerodynamic measures track the pressures and flows in the larynx that occur during voicing. Because the vocal folds act as a valve, pressure is created under the closed glottis and airflow is released through the vibrating folds. Aerodynamic measures are thus able to provide indirect information about the respiratory and laryngeal systems.[60] Validity and reliability for aerodynamic measures is common in the literature (eg, references 61–64).

### *Subglottic pressure*

Estimated subglottic pressure is taken from a pressure-sensing tube placed in the mouth during production of a pressure consonant, typically /p/. Because the pressure in the lungs rapidly is transmitted to the lips, a useful estimate of the subglottic pressure can be obtained.[62] Subglottic pressure has established criterion validity compared with tracheal puncture, and direct measures of subglottic pressure and indirect estimated subglottic pressure at the lips have been demonstrated to have comparable results.[62,65,66]

### *Mean airflow rate*

The mean airflow rate during voicing refers to the average rate of airflow in liters per second during a given production.[62] This measure is commonly taken using an anesthesia-type mask placed over the nose and mouth so that oral airflow during vowel production is passed through a pneumotachometer, which senses pressure changes and mathematically converts these into airflow rates.

## Visual perceptual measures

### *Laryngeal videostroboscopy and high-speed digital imaging*

These imaging tools are commonly used methods to view vocal fold vibration. These tools permit direct visualization of the vibrating vocal folds, allowing detailed assessment of laryngeal structure and function.[67] Because of the large number of vocal fold vibration parameters and somewhat subjective nature of interpreting visual examinations, these measures do hold some reliability concerns. Some of these concerns are examiner bias, clinician training, and lack of standardization in rating parameters. Quantification of imaging parameters has not gained universal acceptance clinically because of its cumbersome nature.[68,69] There are, however, many rating scales that may be used to guide interpretation of imaging parameters.[67,70–72] There are some reliability data in interpretation of imaging parameters, which supports its use. Intrajudge reliability for overall ratings has been demonstrated to range from 73% to 100%, and interjudge reliability was also demonstrated to be strong. More detailed information can be found in the literature (eg, references 70–72).

# METHODS

An electronic database search was carried out using PubMed and EBSCOHost. The search was performed between December 1, 2015 and June 1, 2017. The inclusion criteria consisted of papers in English that were related to any type of voice rehabilitation program that employed VFEs in human subjects. There were no limits set on the time period for the literature search. Papers in other languages and those without peer-review evaluations, abstracts, or incomplete texts were excluded. Search terms used included vocal function exercises, voice therapy, voice management, voice treatment, voice rehabilitation, and voice disorders. There were no restrictions set on types of study designs included. Two reviewers, the first and second author, assessed each abstract for potential inclusion and reached a consensus for the final articles to be included in the review.

## Eligibility

Peer-reviewed studies that fulfilled specific PICO (Patient, Intervention, Comparison, Outcomes) requirements were included in the review. Specific PICO requirements are listed as follows:

**P**atient/**P**opulation: Non–treatment-seeking subjects with normal voice and no identifiable vocal fold pathology, professional voice users, treatment-seeking patient populations, all age groups included
**I**ntervention: Vocal function exercises (VFEs)
**C**omparison: Voice therapy interventions other than VFEs
**O**utcomes: Physiologic voice parameters (acoustics and aerodynamics), V-RQOL, auditory-perceptual voice characteristics, laryngeal imaging findings
**Limits used**: Studies published in English, human subject studies
**Search string**: Search string details are included in Table 1
**Appraisal:** STROBE (Strengthening the reporting of observational studies in epidemiology)[73] checklist items were utilized to appraise study rigor. STROBE checklists were chosen as they include appraisal information for varying study designs. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were utilized to guide the reporting of the present systematic review.

# RESULTS AND DISCUSSION

The search strategy yielded 83 hits across all databases. Of these, 56 articles were rejected after the specific inclusion criteria of VFEs being an intervention method was applied. Of the 27 articles left, one article was excluded as it focused on the utilization of VFEs for the sole purpose of pitch modification in the transgender population.[4] As the study did not focus on enhancement of voice parameters, the study was excluded. Two single case studies by Radhakrishnan et al[14,28] were excluded because the authors used a modified VFE protocol and not the conventional VFE protocol. Case studies by Kumar et al were excluded because VFEs were not implemented as a primary treatment. VFEs were administered in conjunction with Lessac-Madsen Resonant Voice Therapy and specifics of the therapy protocol were not described in the article.[27] The last study to be excluded was an original research study by Sayles. However, this was a Master's thesis available online and not a peer-reviewed paper.[74] Sayles investigated the efficacy of VFEs on aerodynamic parameters in children receiving voice lessons. Based on inclusion or exclusion criteria, 21 studies were included in the final review (Figure 1).

## Evidence levels for included studies

The levels of evidence for each of the 21 studies included in the review were appraised using the standards set by the American Speech-Language and Hearing Association.[75] The levels of evidence for each study are listed in Table 2. The majority of the studies included were level I studies, followed by levels III and II. Study designs that were well-designed randomized clinical trials were classified as level Ib studies. Well-designed nonrandomized studies that included control groups were classified as level IIa. Participant numbers ranged from a minimum of 1[12,17] to a maximum of 134 participants.[20] The ages of

**TABLE 1.**
**Free Text Words or Search Terms and MeSH String (PubMed)**

| Type of Search Terms | Literature Database | Search Terms | Limits | Number of Abstracts Identified |
|---|---|---|---|---|
| Free text words | PubMed | (vocal function exercises) AND (voice therapy) AND (voice treatment) AND (voice rehabilitation) AND (voice disorders) | Humans | 83 |
| MeSH or thesaurus terms | PubMed search details | Search details: (vocal[All Fields] AND ("physiology"[Subheading] OR "physiology"[All Fields] OR "function"[All Fields] OR "physiology"[MeSH Terms] OR "function"[All Fields]) AND ("exercise"[MeSH Terms] OR "exercise"[All Fields] OR "exercises"[All Fields] OR "exercise therapy"[MeSH Terms] OR ("exercise"[All Fields] AND "therapy"[All Fields]) OR "exercise therapy"[All Fields])) AND (("voice"[MeSH Terms] OR "voice"[All Fields]) AND ("therapy"[Subheading] OR "therapy"[All Fields] OR "therapeutics"[MeSH Terms] OR "therapeutics"[All Fields])) AND (("voice"[MeSH Terms] OR "voice"[All Fields]) AND ("therapy"[Subheading] OR "therapy"[All Fields] OR "treatment"[All Fields] OR "therapeutics"[MeSH Terms] OR "therapeutics"[All Fields])) AND (("voice"[MeSH Terms] OR "voice"[All Fields]) AND ("rehabilitation"[Subheading] OR "rehabilitation"[All Fields] OR "rehabilitation"[MeSH Terms])) AND ("voice disorders"[MeSH Terms] OR ("voice"[All Fields] AND "disorders"[All Fields]) OR "voice disorders"[All Fields]) | | |
| Free text words | EBSCOhost | (vocal function exercises) OR (voice therapy) OR (voice treatment) OR (voice rehabilitation) AND (voice disorders) | English, abstracts available | 28 |

participants across studies ranged from 22 to 90 years. Studies focused on varying populations, including non–treatment-seeking subjects under the age of 60 years with normal voice,[3,18,21] the aging population (ages over 60),[2,6,8,16,19,22] school teachers,[5,10,11,15,20] singers,[7,21] and professional voice users.[13]

At least one group of participants across all studies received VFEs. Comparison therapy methods included vocal hygiene (VH), flow resistant tube exercises (FRT), vocal amplification (VA), phonation resistance training exercise (PhoRTE), control (no treatment), and a partial VFE program. Patients received therapy across a range of periods ranging from 1 day to 12 weeks of direct contact sessions. The longest duration of voice therapy was 3 months;[6,17] however, the specific direct therapy contact time was not specified. On average, participants attended 4–6 weeks of voice therapy involving VFEs. Voice outcomes included fell into at least one of the five domains of voice assessment described earlier. Effect sizes were calculated across common voice outcomes over two or more studies. Effect sizes were calculated using Hedge's g correction, which is a less biased effect-size calculation that is more useful for smaller sample sizes. Effect sizes are listed in Table 3. Figure 2 contains a Forest plot with effect sizes classified by outcomes. The Forest plot contains effect sizes for the most commonly used measures across studies.

### *Patient self-report measures*

Six studies included in the review utilized the VHI,[8,9,14–16,26] four studies utilized the V-RQOL questionnaire,[2,5,13,22] one study utilized the Voice Activity Participation Profile (VAPP),[20] one study utilized the Voice Symptom Scale (VoiSS),[5] and three studies utilized nonstandardized scales to assess self-perception related to severity of one's voice problem[10,11,22] and the participants' perceptions on effectiveness of the VFE program in improving their voice quality. Improvement in voice-related self-report measures was reported across all participant groups that performed VFEs. A worsening of voice-related self-report measures was not reported across any study.

On further evaluation of the included studies, Berg et al,[2] Kaneko et al,[8] Gillivan-Murphy et al,[5] and Roy et al[15] showed a statistically significant improvement in self-report measures compared with no treatment groups. Studies by Berg et al and Kaneko et al were both retrospective case-control studies and included participants over the age of 60 years.[2,8] Participants in

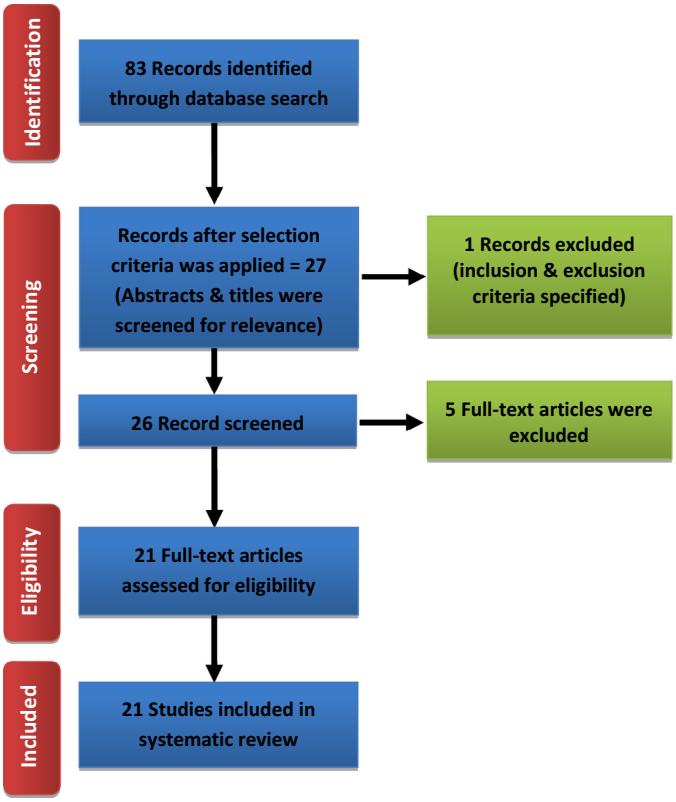


**FIGURE 1.** Flowchart for inclusion or exclusion of articles.

the prospective randomized trial by Roy et al[15] were schoolteachers and the study included a vocal hygiene group in addition to a no treatment group. Gillivan-Murphy et al,[5] who also performed a prospective randomized clinical trial, reported a greater improvement in V-RQOL scores in the VFE+ hygiene group compared with a no treatment group; however, these changes were not statistically significant. The study of Gillivan-Murphy et al, however, did show a greater statistically significant improvement in VoiSS scores compared with the no treatment group.

Comparison voice therapy methods were used by Pedrosa et al (comprehensive voice rehabilitation program),[13] Roy et al (vocal hygiene, no treatment),[15] Kapsner-Smith et al (flow resistant tubes),[9] Nguyen and Kenny (partial VFE program),[10] Teixeira et al (VA),[20] and Ziegler et al (phonation resistance training).[22] Participants across all of these studies were recruited prospectively and therapeutic group assignment was randomized, elevating the level of evidence for this group of studies. The study results from Roy et al[15] and Teixeira and Behlau[20] showed statistically significant improvement in self-report measures. Significant differences in self-report scores were not noted in results reported by Pedrosa, Nguyen, Ziegler and Kapsner-Smith[9,10,13,22] Pedrosa, Ziegler, and Kapsner-Smith[9,13,22] utilized standardized tools for patient self-report, whereas the study by Nguyen and Kenny[10] used a nonstandardized self-report scale. Results from Teixeira and Behlau[20] showed an improvement in self-report measures across all domains in the VFE group compared with the VA group. The VA group showed an improvement in scores in some domains of the VAPP; however, it did not show an improvement in daily and social communication scores.[20]

Effect sizes (Table 3) for patient self-report measures on the VHI and V-RQOL ranged from −1.35 to −0.42 for the V-RQOL and 0.54 to 0.89 for the VHI. The smallest effect was seen in the study by Pedrosa et al[13] at −0.42 (medium effect) and the largest effect was seen in the study by Gillivan-Murphy et al[5] at 0.89.

### *Auditory-perceptual voice measures*

Of the 20 studies included, 5 studies utilized auditory-perceptual analysis as part of the voice assessment protocol.[8,9,12,13,20] Kapsner-Smith et al,[9] Patel et al,[12] and Teixeira and Behlau[20] utilized the CAPE-V.[44] Kaneko et al[8] and Jafari et al[26] utilized the GRBAS scale to perform a perceptual analysis of voice. Both the CAPE-V[44] and GRBAS[45] are validated tools for perceptual assessment of voice. It is not clear whether the tools used by Pedrosa et al for perceptual analysis were validated.[13] Perceptual analyses across all studies were performed by experienced blinded listeners who were experts in the field of voice disorders. A statistically significant improvement was noted in perceptual characteristics of voice across all studies[8,9,12,20,26] except Pedrosa et al.[13] A worsening of voice quality was not reported in any of the studies included in the review. In the study by Kapsner-Smith et al,[9] a significant reduction of roughness was noted in the FRT group compared with the VFE group. In the study by Texeira and Behlau,[20] a significant reduction in overall severity of dysphonia was noted in the VFE group compared with the control group and VA group.

### *Acoustic analysis*

Eight studies utilized acoustic analysis as part of their voice assessment battery.[3,7,8,10–12,20] Commonly used measures included jitter,[8,10,12,20] shimmer,[8,10,12,20] noise to harmonics ratio,[8,10,12,20] and maximum phonation frequency range.[3,11,21] Lesser utilized measures such as long-term average spectrum measures were used in a study by Guzman et al.[7] Improvement in acoustic parameters was noted in all studies except for one. The study by Patel et al,[12] which was a case study, showed normal acoustic parameters before and after VFEs. The study participant had a contact granuloma and not a vocal fold cover pathology which would explain the obtained normative values on the pre- and post-therapy assessment. Study findings from Kaneko et al,[8] Tay et al, Texeira et al,[20,23] and Nguyen and Kenny,[10] showed an improvement in perturbation and noise to harmonics parameters as a result of VFEs application. Findings from Ellis and Beltyukova[3] showed an improvement in maximum phonation frequency range in the monitored VFE group. However, Pasa et al[11] and Wrycza Sabol et al[21] did not show any differences in frequency range across the three comparison groups, one of which was VFEs. Two studies calculated noise in the voice signal. One study calculated harmonics to noise,[10] whereas the other study calculated noise to harmonics[20] precluding direct comparison.

The effect sizes (Table 3) for jitter percentage ranged from small to medium. The smallest effect size (0.05) was seen in the study by Sauder et al,[16] and the largest effect was seen in the study by Kaneko et al[8] at −0.74. Effect sizes for Shimmer ranged from small to medium. The smallest effect size was seen in the

**TABLE 2.**
**Levels of Evidence†**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Berg et al, 2008 | Case-control study n = 25, Patients > 60 y of age were chosen. Age range: 62–83 y<br><br>Cases (VFE) = 19, control = 6 | Historical controls, no treatment | III | No | Control: Mean duration: 3.3 mo, Cases: Mean duration: 5.1 mo | No | Patient self-report: V-RQOL | Controls: Pre-V-RQOL: 68 (18.5), Post-V-RQOL: 69 (23.6) Control group (CG), mean improvement score: 0.42, $P = 0.96$ Cases: Pre-V-RQOL: 58 (24.1), Post-V-RQOL: 77 (18.9) Cases, mean improvement score: 19.21, $P < 0.001$* |
| Ellis et al, 2011 | Quasi-experimental design, n = 20. Mean age of participants = 22 years, Group: monitored VFE group. Unmonitored =10, monitored =10 | Unmonitored VFE groups | IIb | No | 28 d (approximately 4 wk) | Yes, for the monitored group only | Aerodynamic measures: Maximum phonation time Acoustic measures: Maximum phonational frequency range Phonation quotient | Intervention was the same. One was monitored, one was not. (1) Monitored-MPT (pre): 21.85 (5.8), post: 36.26 (8.6) , $P = 0.014$* MPFR (pre) 652.1 (188.8), post: 834.5 (192.7), $P = 0.004$* PQ (pre) 160.7 (79.5), post: 92.9 (27.63), $P = 0.226$ (2) Unmonitored-MPT (pre): 20.15 (5.84), post: 26.35(6.85), $P =$ MPFR (pre) 632.89(192.7), post: 680.74 (171.32) PQ (pre) 170.16 (32.9), post: 129 (26.94) |

(*Continued*)

**TABLE 2.**
**(*Continued*)**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Gillivan-Murphy et al, 2006 | Quasi-experimental design. Study performed with teachers. n = 20. Mean age of 38 years. Control = 11, Treatment (VFE) = 9 | Vocal hygiene | IIb | No | 6 wk | No | Patient self-report measures: V-RQOL Voice care knowledge scale: Voice questionnaire developed for the study | (1) Control: V-RQOL—social (pre): 67 (27), post: 76.1(20.3), mean change: 9.1 (13.8),p = 0.057<br>V-RQOL—physical: pre 51.5 (23.7), post: 56.4 (24.7), mean change 4.9 (13), *P* = 0.241<br>V-RQOL—total: pre: 57.7 (22.7), post: 64.3 (21.6), mean change: 6.6 (10.4), *P* = 0.065<br>(2) Treatment: V-RQOL—social: pre:78.5 (23.6), post: 94.4 (5.8), change 16.0 (19.8), *P* = 0.046*<br>V-RQOL—pre:69.0 (16.7), post: 84.3 (8.5), change 15.3 (15.3), *P* = 0.020*<br>V-RQOL—total: pre: 72.8 (18.9), post: 88.3 (6.3), change 15.6 (15.8), *P* = 0.021*<br>(3) Control: VoiSS: impairment: pre: 33.1 (9.6), post: 31.5 (11.7), change: 1.5, *P* = 0.31<br>VoiSS Emotional: pre: 6.4 (4.6), post: 6.9 (5.7), change: 20.5, *P* = 0.631<br>VoiSS Physical: pre:11.4 (4.4), post: 9.8 (4.0), change: 1.5, *P* = 0.094 Total: pre: 50.8 (15.7), post: 48.3 (18.5), change: 2.5, *P* = 0.334<br>(4) Yes treatment: VoiSS: Impairment: pre: 23.9 (9.1), post: 14.8 (6.8), change:9.1, p = 0.007*<br>VoiSS Emotional: pre: 4.4 (5.1), post: 1.7 (3.0), change: 2.8, *P* = 0.035*<br>VoiSS Physical: pre: 8.1 (3.0), post: 5.3 (3.3), change: 2.5, *P* = 0.09 Total: pre: 36.6 (15.5), post: 21.8 (10.6), change: 13.8, *P* = 0.027 |

(*Continued*)

**TABLE 2.**
**(*Continued*)**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Gorman et al, 2008 | Cohort study without CG. Study performed with the aging population. 19 participants. Age range = 60–78 years. N = 19 | No comparison group | III | No | 12 wk | No | Aerodynamic measure: Maximum phonation time | (1) Pre-MPT: Mean –22.02 (6.86), Post-MPT: 37.22 (9.95), $P < 0.001$* |
| Guzman et al, 2012 | Quasi-experimental design. n = 38. Mean age: 34 y. Pop singers. Experimental (VFE) = 20, Control = 18 | CG: Traditional singing warm up exercises | III | No | VFE duration: 15 min. Traditional singing warm-up duration = 15 min | Yes | Acoustic measure: Long-term average spectrum (alpha ratio, L1:Lo ratio, singing power ratio [SPR]) | (1) CG alpha ratio (median and IQR)<br>Pretreatment = –20.19 (2.94), Post-treatment: –18.18 (2.97), $P = 0.0096$*<br>L1:L0 ratio (median and IQR) Pretreatment = –0.38 (5.6), Post-treatment: –0.5 (3.51), $P = 0.39$<br>SPR (median and IQR)<br>Pretreatment: –15.15 (1.91), Post-treatment: –11.21 (4.13)<br>2) VFE group alpha ratio (median and IQR)<br>Pretreatment = –20.64 (2.27), Post-treatment: –19.83 (2.17), $P = 0.0004$*<br>L1:L0 ratio (median and IQR)<br>Pretreatment = 1.17 (4.35), Post-treatment: 1.48 (2.73), $P = 0.4115$<br>SPR (median and IQR)<br>Pretreatment: –13.72 (3.16), Post-treatment: –12.72 (4.47), $P < 0.001$* |

(*Continued*)

**TABLE 2.**
**(*Continued*)**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Kaneko, et al, 2015 | Retrospective case control study, n = 16, 65–81 y. VFE group = 16, Control = 6 | Historical controls, no treatment | III | No | Approximately 8 wk | Yes, through home practice progress sheets | Acoustic measures: Jitter, shimmer Aerodynamic measures: Maximum Phonation Time Patient self report measures: VHI-10 Auditory-perceptual analysis: GRBAS | (1) Control: Pre-MPT: 22.17 (9.24), Post: 17.67 (9.23), $P = 0.053$ Control: Pre-jitter –1.43(0.58), Post-jitter 1.43 (0.66), $P = 0.99$ Control: Pre-shimmer –3.42 (1.19), Post-shimmer –3.3 (0.94), $P = 0.83$ Control: Pre-VHI –10.67 (5.92), Post-VHI –11.67 (3.08), $P = 0.68$ Control: Pre-GRBAS –9.08 (2.5), 9.58 (2.06), $P = 0.11$ Control: Pre-VHI –10.67 (5.92), Post–VHI –11.67 (3.08), $P = 0.68$ 2) VFE: Pre MPT –14.19(5.86), Post –22.25(7.86), $P < 0.001$* VFE: Pre-jitter 1.41(1), Post-jitter 0.89 (0.72), $P = 0.014$* VFE Pre-shimmer –3.8 (2.4), Post-shimmer –3.01 (2.22), $P = 0.1$ VFE: Pre-VHI 18.88 (8.49), Post: 7.56 (4.79), $P < 0.001$* VFE: Pre-GRBAS, 9.44 (1.74), Post: 6.94 (1.18), $P < 0.001$* VFE: Pre-VHI –10–18.8 (8.49), 7.56 (4.79), $P < 0.001$* |
| Kapsner-Smith et al, 2015 | Randomized controlled trial, n = 21. Age range: 32–72 y. VFE = 10, FRT = 11 | Immediate flow resistant tube (FRT) therapy Two CGs: Delayed FRT and delayed VFE group served as no treatment groups | Ib | No | 6 wk, once a week with clinical monitoring | Home practice was unmonitored | Patient self-report measures: VHI Auditory perceptual analysis: CAPE-V | (1) Controls: Pre-VHI: 40 (18.5), Post VHI: 38 (14.4), $P > 0.05$ (not specified) (2) FRT Pre-VHI 39.4 (15), Post-VHI 21.6 (13.5): $P < 0.001$* (3) VFE Pre-VHI 45.2 (17.9), Post-VHI 33.7 (21.2): $P < 0.048$* (4) CAPE-V change coefficient scores were not significant for measures under study |

(*Continued*)

**TABLE 2.**
**(*Continued*)**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Pasa et al, 2007 | Randomized clinical trial, n = 37. Age range = 23–52 y Comparison group = vocal hygiene (VH), no treatment. School teachers. Vocal hygiene = 13, VFE = 10, Control = 14 | Vocal hygiene CG: No treatment | Ib | No | 10 wk | Yes, through home practice progress sheets | Patient questionnaire: voice knowledge (non-standardized tool), perceived degree of benefit questionnaire Aerodynamic measure: Maximum phonation time | (1) VH: Pre-voice knowledge: 24.08 (3.57), Post-voice knowledge: 26.62 (3.43), $P = 0.002$* VFE: Pre-voice knowledge: 24.8 (2.25), Post-voice knowledge: 25.8 (2.57), $P > 0.05$ (not specified) CG: Pre-voice knowledge: 24.43 (2.77), Post-voice knowledge: 24.5 (1.65), *P*-value not specified (2) VH: Prenormalized MPT: –8.6 (4.27), Postnormalized MPT: -6.9 (3.86), *P*-value not specified VFE: Prenormalized MPT: –6.41 (4.54), Postnormalized MPT: –5.74 (3.44), *P*-value not specified CG: Prenormalized MPT: –8.3 (4.32), Postnormalized MPT: –9.06 (6.12), *P*-value not specified |
| Patel et al, 2012 | Case report, n = 1. Age: 51 y | No comparison group | III | No | 6 wk | Yes, through home practice progress sheets | Acoustic measures Aerodynamic measures Auditory-perceptual analysis: CAPE-V high speed digital imaging | N/A |

(*Continued*)

**TABLE 2.**
**(*Continued*)**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Pedrosa et al, 2015 | Randomized clinical trial, n = 80, mean age: 35 y (SD = 10.1), study included patients with functional dysphonia. CVRP = 40, VFE = 40 | Comprehensive voice rehabilitation program (CVRP) | Ib | Yes | 6 wk | Yes, graphic chart to record frequency of exercises | Auditory-perceptual measure: Nonstandardized APE or auditory perceptual evaluation Patient self-report: VHI and V-RQOL | (1) CVRP: Pre-VHI –43 (22.4), Post-VHI–25.4 (20.9) (2) VFE: Pre-VHI –32.52 (20.32), Post-VHI –23.93 (19.09) (3) CVRP: Pre-V-RQOL –64.4 (20.27), Post-V-RQOL 80.75 (18.98) (4) VFE: Pre-V-RQOL –72 (18.7), Post-V-RQOL –82 (15.19) *P*-values are not listed |
| Roy et al, 2001 | Randomized clinical trial. Mean age: 42–44 years, Teachers. VFE = 19, control = 19, Vocal hygiene = 20 | Vocal hygiene CG: No treatment | Ib | No | 6 wk | Yes, through four session over 6 wk | Patient self-report measure: VHI | (1) No treatment pre- and post-VHI changes: $P = 0.233$ (2) Vocal hygiene pre and post VHI changes: $P = 0.918$ (3) VFE group pre- and post-VHI changes: $P < 0.0002$* |
| Wrycza Sabol et al, 1995 | Quasi-experimental design, randomization was not specified, n = 20, singers. Age:21–43 y, VFE = 10, vocal hygiene = 10 | Vocal hygiene | IIa | No | 4 wk | Yes, by maintaining a written log of their VFE maximum phonation times in seconds | Aerodynamic measure: Maximum phonation time Acoustic measure: Jitter | (1) Control: Pre-jitter –0.25 (0.08), Post-jitter –0.18 (0.06), Control: Pre-MPT – 23 (7), Post-MPT: 25 (5) (2) VFE: Pre-jitter –0.24 (0.11), Post-jitter –0.31 (0.11) VFE: Pre-MPT –22 (6), Post-MPT: 31 (11) *P*-values are not specified for pre- and post-changes |

(*Continued*)

**TABLE 2.**
**(*Continued*)**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Sauder et al, 2010 | Cohort study without control, n = 9. Aging population: 67–90 y. VFE = 9 | No comparison group | IIb | No | 6 wk | No | Patient self-report measures: VHI, voice disorder severity<br>Phonatory effort<br>acoustic measures: harmonic to noise ratio, jitter, shimmer<br>Aerodynamic measure: Maximum phonation Time<br>Auditory perceptual ratings: Rainbow passage, Sustained /a/<br>Stroboscopic evaluation: Phase closure | (1) Pre-HNR –0.16 (3.15), Post-HNR 0.14 (2.35), $P = 0.67$<br>(2) Pre-jitter –1.28 (1.23), Post-jitter –1.22 (0.862), $P = 0.77$<br>(3) Pre-shimmer –4.98 (2.27), Post-shimmer –3.84 (2.41), $P = 0.17$<br>(4) Pre-MPT –17.78 (5.35), Post-MPT–18.62 (5.81), $P = 0.68$<br>(5) Pre-VHI –39.11 (21.35), Post-VHI –23.44 (19.54), $P = 0.01$*<br>(6) Pre-Voice disorder severity: 1.7 (0.71), Post-Voice disorder severity: 1.1 (0.93), $P = 0.1+$<br>(7) Pre-phonatory effort: 1.78 (1.09), Post-phonatory effort: 0.89 (0.6), $P = 0.04$*<br>(8) Rainbow Passage—Pre-overall quality: 32.17 (25.26), Post-over all quality: 22.56 (17.83), $P = 0.11$<br>Pre-breathiness: 27(22.04), Post-breathiness: 15.45 (7.6), $P = 0.09+$<br>Pre-strain: 29.39 (25.43), Post-strain: 20.97(19.08), $P = 0.1+$<br>(9) Sustained /a/—Pre-overall quality: 39 (14.78), Post-overall quality: 36 (13.36), $P = 0.21$<br>Pre-breathiness: 23.1 (17.68), Post-breathiness: 24 (14), $P = 0.37$<br>Pre-strain: 39.15 (16.14), Post-strain: 33 (14.29), $P = 0.11$ |

(*Continued*)

**TABLE 2.**
**(*Continued*)**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Sharma et al, 2009 | Case report, n = 1 | No comparison group | III | No | 3 mo | No | Aerodynamic measures: Maximum phonation time | Pre-therapy MPT: 7.8 s, Post-therapy MPT: 19.8 s |
| Stemple et al, 1994 | Randomized clinical trial, age range: 21–25 y, n = 35, comparison groups = control and placebo. | No treatment, Placebo | Ib | No | 4 wk | Yes, subjects logged home practice phonation times on the VFEs | Aerodynamic measure: Maximum phonation time, flow rate, flow volume<br>Acoustic measure: Jitter | (1) Control: Pre-MPT –20 (5), Post-MPT: 20 (5)<br>Placebo: Pre-MPT: 21 (6), Post-MPT: 19 (7)<br>VFE: Pre-MPT: 24 (6), Post-MPT: 27 (7)<br>(2) Control: Pre-jitter –0.26 (0.14), Post-jitter –0.38 (0.19)<br>Placebo: Pre-jitter –0.35 (0.21), Post-jitter –0.36 (0.21)<br>VFE: Pre jitter 0.4 (0.19), Post jitter –0.59 (0.16) |
| Tanner et al, 2010 | Case study, Twin study, n = 2, aging voice population, Age: 79 y | No comparison group | III | No | 4 sessions, frequency of sessions was not specified | No | Auditory-perceptual evaluation: Tool not specified<br>Acoustic measures<br>Aerodynamic measures<br>EMG | n/a Case report |

(*Continued*)

**TABLE 2.**
**(*Continued*)**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Tay et al, 2012 | Quasi experimental design, n = 22, age range: 68–83 years, aging choral singers. VFE = 11, Control = 11 | CG: No treatment | IIa | No | 5 weeks | No | Acoustic measures: Jitter, shimmer Aerodynamic measures: Maximum phonation time Auditory perceptual measures: breathiness, roughness, Strain | (1) Control: Pre-MPT –17.87 (4.38), Post-MPT: 17.93 (4.98) VFE: Pre-MPT –17.86 (6.78), 21.96 (8.32) (2) Control: Pre-jitter –1.62 (1.4), Post-jitter –0.73 (0.39) VFE: Pre-jitter –1.39 (0.77), Post-jitter –1.03 (0.55) 3) Control: Pre-shimmer –2.58 (1.18), Post-jitter –2.22 (1.06) VFE: Pre-shimmer –3.37 (1.43), Post-jitter –1.92 (1.13) (4) Control: Pre-NHR –0.115 (0.021), Post-NHR –0.102 (0.020) VFE: Pre-NHR –0.134 (0.015), Post-NHR –0.097 (0.022) (5) Control: Pre-breathiness: 2.09 (0.49), Post-breathiness: 2.05 (0.57) VFE: Pre-breathiness: 2.14 (0.87), Post-breathiness: 1.95 (0.72) 6) Control: Pre-roughness: 2.45 (0.61), Post-roughness: 2.23 (0.79) VFE: Pre-roughness: 2.36 (0.45), Post-roughness: 2.0 (0.39) 7) Control: Pre-strain: 2.05 (0.35), Post-strain: 1.86 (0.39) VFE: Pre-strain: 2.09 (0.44), Post-strain: 2.14 (0.64) p-values unavailable |

(*Continued*)

**TABLE 2.**
**(*Continued*)**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Teixeira et al., 2014 | Randomized clinical trial, n = 162, Age: 18–50 years, comparison: voice amplification (VAG), CG. Control (CG) = 54, Voice amplification = 54, VFE = 54 | CG: No treatment, Voice amplification (VAG) | Ib | No | 6 weeks | Yes, through home practice progress sheets | Auditory-perceptual evaluation: Consensus Auditory Perceptual Evaluation –V Laryngeal assessment: Stroboscopy (glottal closure, lesion size, closure and tension) Patient self-report: Voice activity participation profile acoustic measures: jitter, shimmer | (1) Control: Pre-total VAPP –40.71 (47.2), Post-Total VAPP: 49.01 (58.71) VAG: Pre-total VAPP –47.38 (48.1), Post-Total VAPP: 35.42 (39.71)<br>VFE: Pre-total VAPP –67.59 (63.6), Post-Total VAPP: 37.92 (55.0)<br>(2) Control: Pre-shimmer –5.7 (2.3), Post-shimmer: 5.22 (1.9) VAG: Pre-Shimmer –4.72 (1.7), Post-shimmer: 4.24 (1.4) VFE: Pre-shimmer –5.09 (2.1), Post-shimmer: 3.86 (1.6) 7)<br>(3) Control: Pre-jitter –1.07 (0.7), Post-jitter: 0.83 (0.5) VAG: Pre-jitter –0.69 (0.3), Post-jitter: 0.61 (0.3) VFE : Pre-jitter –0.89 (0.7), Post-jitter: 0.62 (0.4)<br>(4) Control: Pre NHR –0.16 (0.7), Post-jitter: 0.14 (0.5) VAG: Pre-NHR –0.13 (0.3), Post-NHR: 0.13 (0.3) VFE: Pre-NHR –0.13 (0.7), Post-jitter: 0.12 (0.4)<br>Pre- and post-therapy intervention CAPE-V scores are not available |

(*Continued*)

**TABLE 2.**
**(*Continued*)**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Ziegler et al, 2014 | Randomized clinical trial, n = 20. Age: 60–91 y. VFE = 6, PhoRTE = 6, Control = 8 | Phonation resistance training exercise (PhoRTE), Control: No treatment | Ib | No | 4 weeks | Yes, through home practice progress sheets | Patient self-report: V-RQOL, perceived phonatory effort (PPE) | (1) Control: Pre-V-RQOL: 87.5 (7.9), Post-V-RQOL: 91.5 (5.8), $P = 0.195$<br>PhoRTE: Pre-V-RQOL: 88.5 (10.4), Post-V-RQOL: 95.0 (5.9), $P = 0.54$*<br>VFE: PhoRTE: Pre-V-RQOL: 80.8 (12.3), Post-V-RQOL: 87.5 (10.2), $P = 0.054$*<br>(2) Control: Pre-PPE: 101 (16), Post-PPE: 103 (13), $P = 0.374$<br>PhoRTE: Pre-PPE: 144 (51.8), Post-PPE: 102 (35.6), $P = 0.77$*<br>VFE: PhoRTE: Pre-PPE: 142.5 (41.7), Post-PPE: 109.2 (14.3), $P = 0.121$ |
| Nguyen et al, 2009 | Randomized clinical trial, primary school teachers, n = 40. Age range = 22–54 y. Full VFE protocol = 22, Partial protocol = 18 | Partial VFE protocol | Ib | No | 4 wk | Yes, through home practice progress sheets | Acoustic measures: Jitter, shimmer, HNR<br>Auditory-perceptual evaluation<br>Patient self-report measures | (1) Full exercise group: Pre-jitter –0.599, Post-jitter –0.356, $P < 0.001$*<br>Partial exercise group: Pre-jitter –0.563, Post-jitter –0.529, $P = 0.511$<br>(2) Full exercise group: Pre-shimmer –4.343, Post-shimmer –2.935, $P < 0.001$*<br>Partial exercise group: Pre-shimmer –4.086, Post-shimmer –3.942, $P = 0.748$<br>(3) Full exercise group: Pre-HNR –18.6, Post-HNR –22.4, $P < 0.001$*<br>Partial exercise group: Pre-HNR –18.9, Post-HNR –19.6, $P = 0.206$<br>(4) Full exercise group: Pre-perceptual scores –2.82, Post-perceptual scores –2.41, $P = < 0.001$*<br>Partial exercise group: Pre-perceptual score –2.87, Post-perceptual score –2.72, $P = 0.068$ |

(*Continued*)

**TABLE 2.**
**(*Continued*)**

| Author and Year | Study Design and Population | Comparison Group/ Treatment | Level of Evidence | Power Analysis Included | Treatment Duration | Adherence Monitored | Outcome Measures | SDs, Means, and *P*-Values |
|---|---|---|---|---|---|---|---|---|
| Jafari et al, 2016 | Cohort study. Native Persian speakers with muscle tension dysphonia, n = 15. Age range: 24–62 y | No comparison group | IIb | No | 6 wk | No | Patient self-assessment: VHI<br>Auditory-perceptual assessment: GRBAS | (1) Pre-VHI total: 43.4 (19.2), Post-VHI total: 24.4 (18.9), $P = 0.02$*<br>Pre-VHI Physical: 19.4 (9.2), Post-VHI Physical: 11.5 (8.1), $P = 0.006$*<br>Pre-VHI Functional: 17.8 (8.4), Post-VHI Functional: 8.6 (6.3), $P = 0.04$*<br>Pre-VHI Emotional: 6.2 (5.9), Post-VHI Emotional: 4.3 (4.4), $P = 0.004$*<br>(2) Pre-overall grade of dysphonia (G): 2.1 (0.8), Post-overall grade of dysphonia (G): 0.5 (0.6), $P = 0.006$*<br>Pre-Roughness (R): 1.8 (0.6), Post-Roughness (R): 0.7 (0.5), $P = 0.01$*<br>Pre-Breathiness (B): 1.2 (0.3), Post-Breathiness (B): 0.4 (0.2), $P = 0.02$*<br>Pre-Asthenia (A): 1.1 (0.5), Post-Asthenia: 0.3 (0.4), $P = 0.04$*<br>Pre-Strain (S): 2.3 (0.9), Post-Strain (S): 0.9 (0.8), $P = 0.004$* |

* Significance level at $P < 0.05$.
† Significance level at $P < 0.1$.
*Abbreviations:* CAPE-V, Consensus Auditory Perceptual Evaluation of Voice; GRBAS, grade, roughness, breathiness, asthenia, and strain; HNR, Harmonic-to-noise ratio; IQR, Interquartile Range; MPT, maximum phonation time; NHR, Noise-Harmonic Ratio; VAPP, Voice Activity Participation Profile; VFE, vocal function exercises; VHI, Voice Handicap Index; VoiSS, Voice Symptom Scale; V-RQOL, Voice-Related Quality of Life.

**TABLE 3.**
**Effect Sizes for Domains of Voice Assessment**

| Parameters | Studies | Effect Size (Hedge d) | Interpretation |
|---|---|---|---|
| Voice Related Quality of Life | Berg et al[2] | 0.8 | Medium to large |
| | Ziegler et al[22] | 0.54 | |
| | Pedrosa et al[13] | 0.56 | |
| | Gillivan-Murphy et al[5] | 0.89 | |
| | Overall V-RQOL | 0.66 | |
| Voice Handicap Index | Jafari et al[26] | –0.99 | Medium to large |
| | Sauder et al[16] | –0.59 | |
| | Pedrosa et al[13] | –0.42 | |
| | Kaneko et al[8] | –1.35 | |
| | Kapsner-Smith et al[9] | –0.53 | |
| | Overall VHI | –0.52 | |
| Maximum phonation time | Sauder et al[16] | 0.11 | Small to large |
| | Pasa et al[11] | 0.3 | |
| | Kaneko et al[8] | 1.00 | |
| | Teixeira and Behlau[20] | 0.49 | |
| | Stemple et al[18] | 0.43 | |
| | Wrycza Sabol et al[22] | 0.85 | |
| | Ellis and Beltyukova[3] | 0.87 | |
| | Gorman et al[6] | 1.55 | |
| | Overall MPT | 0.55 | |
| Jitter | Sauder et al[16] | –0.06 | Small to medium |
| | Teixeira and Behlau[20] | –0.46 | |
| | Wrycza Sabol et al[21] | –0.61 | |
| | Kaneko et al[8] | –0.74 | |
| Shimmer | Wrycza Sabol et al[21] | –0.77 | Small to medium |
| | Teixeira and Behlau[20] | –0.26 | |
| | Kaneko et al[8] | –0.14 | |
| | Sauder et al[16] | –0.49 | |

*Abbreviations:* VHI, Voice Handicap Index; V-RQOL, Voice-Related Quality of Life.

study by Kaneko et al[8] at –0.14. The largest effect was seen in the study by Teixeira and Behlau[20] at –0.77.

### *Aerodynamic measures*

Eight studies utilized aerodynamic analysis as part of their voice assessment battery.[6,8,12,17–19,21] Studies that used maximum phonation time (MPT) as an outcome measure have been included in this section as well. MPT emerged as the most widely used outcome measure across studies. Worsening of aerodynamic measures was not noted in any of the studies. Six out of eight studies reported an improvement in MPTs and aerodynamic measures as a result of VFEs.[6,8,17,18,21] Majority of these study designs included control groups;[8,18,21] one study was a cohort study without controls[6] and two studies were single case studies.[12,17] The only study that did not show a significant change in MPTs as a result of VFEs was a twin study by Tanner et al.[19] Other aerodynamic measures included were airflow rates[12,18,21] and inverse filter measures.[6] The study by Gorman et al[6] showed an improvement in DC flow measures in individuals over the age of 60 years with a diagnosis of presbylarynx. Before therapy, all participants showed DC flow measures consistent with glottic insufficiency observed on their pretherapy stroboscopic examination. At the end of 12 weeks, a significant improvement in DC flow and subglottic pressure measures was noted as a result of VFEs indicative of an improvement in glottic efficiency. Decrease in airflow rates was also noted by Stemple et al[18] and Wrycza Sabol et al[21] as a result of VFEs. Both studies showed a decrease in airflow rates in the VFE group in comparison to placebo therapy and vocal hygiene. The case study by Patel et al[12] also showed a decrease in airflow measures; however, the study participant had normal values on airflow measures during the pre- and post-therapy assessments.

The effect sizes (Table 3) for MPT ranged from small to large, 0.11 to 1.55. Three out of nine showed a small effect size.[11,16,19] The smallest effect size was noted in Sauder et al,[16] at 0.11. However, the study by Tanner et al was a twin study (n = 2). Effect sizes are often unreliable with a small sample size.[76] The largest effect was seen in the study by Gorman et al,[6] at 1.55.

### *Laryngeal imaging*

Four studies utilized laryngeal imaging as an outcome measure. Three studies utilized laryngeal videostroboscopic measures[8,13,16] and one study utilized high-speed laryngeal imaging.[12] No adverse laryngeal findings were reported in any of the studies as a result of VFEs. Study findings from Kaneko et al[8] and Sauder et al[16] showed an improvement in glottic closure as a result of VFEs. The study by Kaneko et al[8] investigated three vibratory parameters which included normalized wave mucosal amplitude

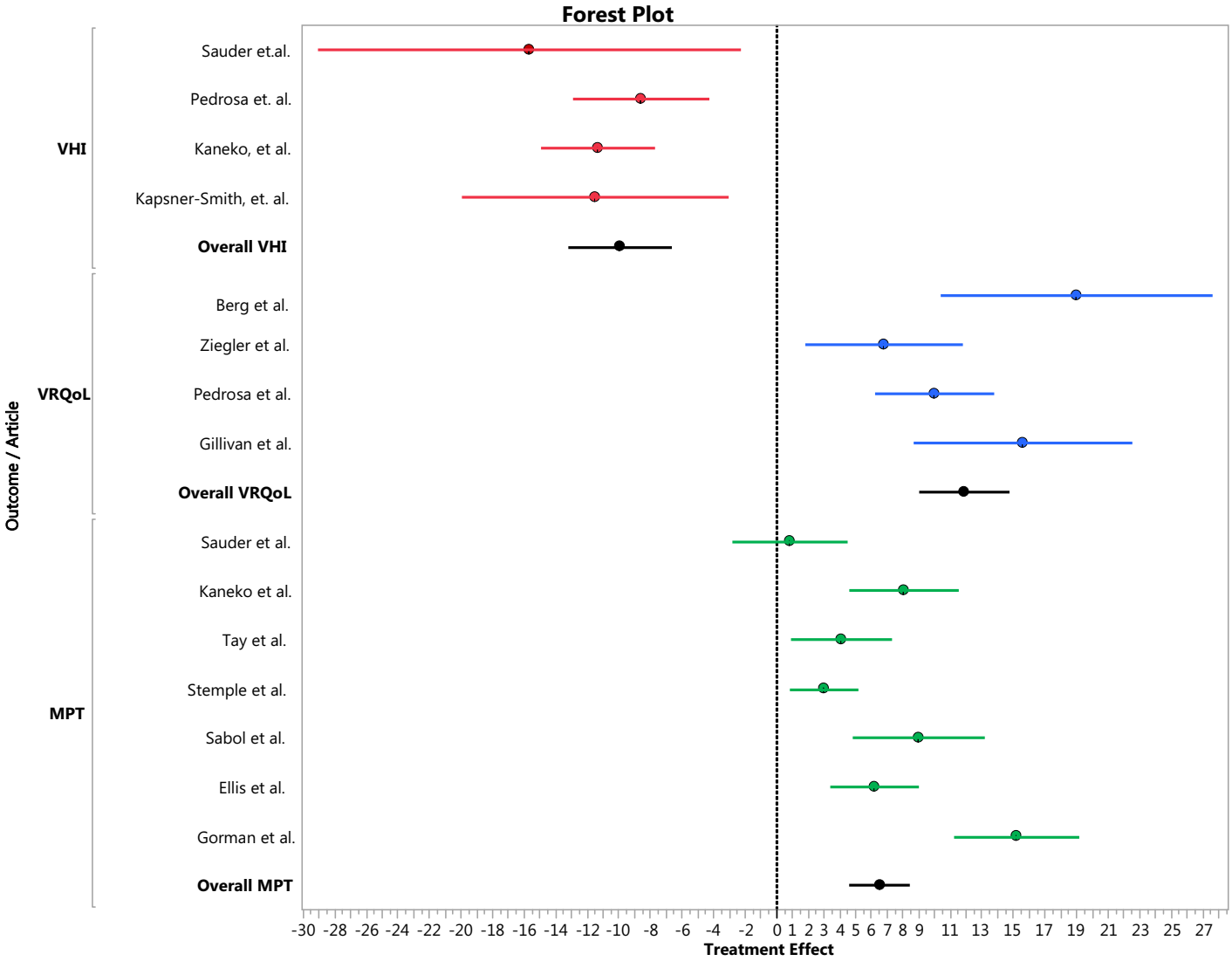


**FIGURE 2.** Forest plot representing effect sizes by specific study and outcome measure.

(NMWA), normalized glottal gap (NGG) and bowing index (BI). The authors found an improvement in NMWA and NGG following an 8-week VFE program. Participants between the ages of 65 and 80 years were included in the study, which would explain the need for a bowing index. Kaneko et al,[8] however, did not observe a difference in the bowing index as a result of VFEs. The study by Pedrosa et al[13] showed an improvement in the comparison group (comprehensive voice rehabilitation program) over VFEs in terms of laryngeal pattern (LP). LP was judged by an otolaryngologist who was blinded to the patient's history and demographic information. LP was judged on a 100-mm visual analog scale in terms of the degree of abnormality of glottic closure pattern, presence or absence of a lesion, and the degree of supraglottic activity. It is unclear as to which LP parameters showed the greatest degree of improvement based on the results presented. Patel et al[12] utilized both high-speed laryngeal imaging (HSDI) and laryngeal stroboscopy to judge improvement in vocal fold vibratory features in a single participant with a vocal process granuloma. The authors reported changes in several HSDI features including improved voice onset time, maximum amplitude, speed quotient, open quotient, peak closing velocity, peak-to-average closing velocity and peak to average opening velocity.[12] On both high speed and stroboscopy, improved closure of the membranous vocal folds was observed during the post-VFE assessment.

### *Tracking adherence*

Thirteen out of 21 studies documented monitoring adherence to therapy techniques under study.[3,7,8,10–13,15,18,20–22,26] Adherence for VFEs was monitored through home practice progress sheets wherein study participants were required to log in phonation times on the VFEs. In the study by Ellis and Beltyukova,[3] participants in the monitored VFE group were required to provide video or audio recordings of the exercises completed on a home practice basis.

## SUMMARY AND CONCLUSIONS

Since the mid-1990s, there has been a steady interest in the efficacy of VFEs for improving vocal function; thus, a thorough review was warranted. The present review did not exclude studies based on study designs. The majority of studies included in the review showed a high study rigor with many assessed as level I or level II. Thirteen studies included some form of a control or no treatment group and nine studies compared VFEs with another voice therapy method. Studies that included a control group or a therapy comparison group utilized randomization for participant group assignments. Effect sizes for the studies under review ranged from small to large. However, the majority of the studies showed medium to large effect sizes. Studies that were designed as randomized clinical trials were all single blinded, further increasing their level of evidence. However, even study designs that are appraised with lower levels of evidence were still executed with a high level of rigor. The case control and case studies included in this review also involved some form of blinding for the auditory perceptual and stroboscopic parameters. To observe such a high level of rigor in the field of voice disorders is impressive and makes one optimistic for the level of research one can expect for the future.

From the appraised studies, it is apparent that VFEs are efficacious in enhancing vocal function in individuals with disordered voices, individuals over the age of 60 years diagnosed with presbylaryngeus, and professional voice users. In individuals with normal voices and elite voice users such as singers, VFEs were effective in enhancing existent physiologic voice parameters such as pitch range, airflow rate, MPT, and dynamic range.

In terms of the voice assessment parameters, patient self-report was the most widely used assessment parameter influencing the outcome of this review. It is interesting that patient-reported measures represented the strongest evidence for the effectiveness of VFEs according to the effect sizes. Other measures that are typically considered more "objective" demonstrated lesser effect. One reason for this may be that the objective measures only captured portions of the many interacting components of voice production. Because vocal output is an emergent behavior resulting from the nonlinear interactions of multiple variables within and among the respiratory, phonatory, and resonance subsystems, measurement of a single variable is likely not a strong representation of the total change in vocal output resulting from VFEs. Self-report measures of vocal output represent the conglomeration of many variables that result in overall vocal function. Future studies should consider multimeasurement methodologies that capture the interactive nature of the subsystem triad, which may better represent the overall changes in vocal output resulting from the VFE program. From the studies included, VFEs appear to influence change in multiple voice parameters across the five domains of voice assessment.

Compliance also appeared to be a factor in the magnitude of improvement noted as a result of VFEs. In the study by Ellis et al, the group that was not monitored for VFEs did poorly compared with the monitored VFE group. In general, the success of VFE was attributed to the monitoring required as part of the therapy program.

In summary, there is moderate-to-strong evidence based on patient self-report and mild-to-moderate evidence based on objective measures to support the use of VFEs to enhance the normal voice and the voices of highly trained singers, remediate disordered voice, and improve the aging voice. It is possible that VFEs have been widely studied because it is a highly prescriptive exercise program, enhancing documentation of progress and changes in vocal function. The present systematic review strengthens the external validity of these exercises as improvements in diverse populations have been observed. The results of this systematic review indicate that VFEs demonstrate both efficacy and effectiveness as a tool for the remediation of voice disorders.

## LIMITATIONS

The analysis of bias for the present review was limited due to small sample sizes and multiple outcome variables used across studies. Funnel plots for three primarily used outcome measures were constructed; however, they cannot be reliably interpreted due to the small number of studies that were included for each outcome. Therefore, the possibility of publication bias exists and should be considered.

## REFERENCES


1. Stemple JC, Glaze L, Gerdeman B. *Clinical Voice Pathology: Theory and Management*. Boston, MA: Cengage Learning; 2000.
2. Berg EE, Hapner E, Klein A, et al. Voice therapy improves quality of life in age-related dysphonia: a case-control study. *J Voice*. 2008;22:70–74.
3. Ellis LW, Beltyukova SA. Effects of compliance monitoring of vocal function exercises on voice outcome measures for normal voice. *Percept Mot Skills*. 2011;112:729–736.
4. Gelfer MP, Van Dong BR. A preliminary study on the use of vocal function exercises to improve voice in male-to-female transgender clients. *J Voice*. 2013;27:321–334.
5. Gillivan-Murphy P, Drinnan MJ, O'Dwyer TP, et al. The effectiveness of a voice treatment approach for teachers with self-reported voice problems. *J Voice*. 2006;20:423–431.
6. Gorman S, Weinrich B, Lee L, et al. Aerodynamic changes as a result of vocal function exercises in elderly men. *Laryngoscope*. 2008;118:1900–1903.
7. Guzman M, Angulo M, Munoz D, et al. Effect on long-term average spectrum of pop singers' vocal warm-up with vocal function exercises. *Int J Speech Lang Pathol*. 2013;15:127–135.
8. Kaneko M, Hirano S, Tateya I, et al. Multidimensional analysis on the effect of vocal function exercises on aged vocal fold atrophy. *J Voice*. 2015;doi:10.1016/j.jvoice.2014.10.017.
9. Kapsner-Smith MR, Hunter EJ, Kirkham K, et al. A randomized controlled trial of two semi-occluded vocal tract voice therapy protocols. *J Speech Lang Hear Res*. 2015;58:535–549.
10. Nguyen DD, Kenny DT. Randomized controlled trial of vocal function exercises on muscle tension dysphonia in Vietnamese female teachers. *J Otolaryngol Head Neck Surg*. 2009;38:261–278.
11. Pasa G, Oates J, Dacakis G. The relative effectiveness of vocal hygiene training and vocal function exercises in preventing voice disorders in primary school teachers. *Logoped Phoniatr Vocol*. 2007;32:128–140.
12. Patel RR, Pickering J, Stemple J, et al. A case report in changes in phonatory physiology following voice therapy: application of high-speed imaging. *J Voice*. 2012;26:734–741.
13. Pedrosa V, Pontes A, Pontes P, et al. The effectiveness of the comprehensive voice rehabilitation program compared with the vocal function exercises method in behavioral dysphonia: a randomized clinical trial. *J Voice*. 2015;doi:10.1016/j.jvoice.2015.03.013.
14. Radhakrishnan N, Scheidt T. Modified vocal function exercises: a case report. *Logoped Phoniatr Vocol*. 2012;37:123–126.

15. Roy N, Gray SD, Simon M, et al. An evaluation of the effects of two treatment approaches for teachers with voice disorders: a prospective randomized clinical trial. *J Speech Lang Hear Res*. 2001;44:286–296.
16. Sauder C, Roy N, Tanner K, et al. Vocal function exercises for presbylaryngis: a multidimensional assessment of treatment outcomes. *Ann Otol Rhinol Laryngol*. 2010;119:460–467.
17. Sharma N, De M, Martin T, et al. Laryngeal reconstruction following shrapnel injury in a British soldier: case report. *J Laryngol Otol*. 2009;123:253–256.
18. Stemple JC, Lee L, D'Amico B, et al. Efficacy of vocal function exercises as a method of improving voice production. *J Voice*. 1994;8:271–278.
19. Tanner K, Sauder C, Thibeault SL, et al. Vocal fold bowing in elderly male monozygotic twins: a case study. *J Voice*. 2010;24:470–476.
20. Teixeira L, Behlau M. Comparison between vocal function exercises and voice amplification. *J Voice*. 2015;doi:10.1016/j.jvoice.2014.12.012. In press.
21. Wrycza Sabol J, Lee L, Stemple JC. The value of vocal function exercises in the practice regimen of singers. *J Voice*. 1995;9:27–36.
22. Ziegler A, Verdolini Abbott K, Johns M, et al. Preliminary data on two voice therapy interventions in the treatment of presbyphonia. *Laryngoscope*. 2014;124:1869–1876.
23. Tay EY, Phyland DJ, Oates J. The effect of vocal function exercises on the voices of aging community choral singers. *J Voice*. 2012;26.
24. Lim H-J, Kim J-K, Kwon D-H, et al. The effect of vocal function exercise on voice improvement in patients with vocal nodules. *J Korean Soc Speech Sci*. 2009;1:37–42.
25. Roy N, Weinrich B, Gray SD, et al. Three treatments for teachers with voice disorders: a randomized clinical trial. *J Speech Lang Hear Res*. 2003; 46:670–688.
26. Jafari N, Salehi A, Izadi F, et al. Vocal function exercises for muscle tension dysphonia: auditory-perceptual evaluation and self-assessment rating. *J Voice*. 2016;doi:10.1016/j.jvoice.2016.10.009.
27. Kumar R, Sharma P, Vir D, et al. Voice therapy outcomes in type-i sulcus vocalis: case studies. *Int J Clin Exp Otolaryngol*. 2016;2:42–44.
28. Radhakrishnan N, Mathisen D. Idiopathic subglottic stenosis: surgical and therapeutic results in an adult woman. *Int J Clin Exp Otolaryngol*. 2016;2:23–25.
29. Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71–72.
30. Speyer R. Effects of voice therapy: a systematic review. *J Voice*. 2008;22:565–580.
31. Roy N, Merrill RM, Gray SD, et al. Voice disorders in the general population: prevalence, risk factors, and occupational impact. *Laryngoscope*. 2005;115:1988–1995.
32. Roy N, Merrill RM, Thibeault S, et al. Voice disorders in teachers and the general population: effects on work performance, attendance, and future career choices. *J Speech Lang Hear Res*. 2004;47:542–551.
33. Yiu EM. Impact and prevention of voice problems in the teaching profession: embracing the consumers' view. *J Voice*. 2002;16:215–228.
34. Mattiske JA, Oates JM, Greenwood KM. Vocal problems among teachers: a review of prevalence, causes, prevention, and treatment. *J Voice*. 1998;12:489–499.
35. Thibeault SL, Merrill RM, Roy N, et al. Occupational risk factors associated with voice disorders among teachers. *Ann Epidemiol*. 2004;14:786–792.
36. Verdolini K, Ramig LO. Review: occupational risks for voice problems. *Logoped Phoniatr Vocol*. 2001;26:37–46.
37. Cohen SM. Self-reported impact of dysphonia in a primary care population: an epidemiological study. *Laryngoscope*. 2010;120:2022–2032.
38. Stemple JE. *Voice Therapy: Clinical Studies*. 2nd ed. San Diego: Singular Publishing Group; 2000.
39. Thomas L, Stemple J. Voice therapy: does science support the art? *Commun Disord Rev*. 2007;1:49–77.
40. Hirano M. Objective evaluation of the human voice: clinical aspects. *Folia Phoniatr (Basel)*. 1989;41:89–144.
41. Bless DM, Hicks D. Diagnosis and measurement assessing the "WHs" of voice function. In: Brown WS, Vinson BP, Crary MA, eds. *Organic Voice Disorder—Assessment and Treatment.* San Diego, CA: Singular Publishing group; 1996:119–170.
42. ASHA. *Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) ASHA Special Interest Division 3, Voice and Voice Disorders*; 2009.
43. 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.
44. Zraick RI, Kempster GB, Connor NP, et al. Establishing validity of the consensus auditory-perceptual evaluation of voice (CAPE-V). *Am J Speech Lang Pathol*. 2011;20:14–22.
45. Hirano M. Psycho-acoustic evaluation of voice: GRBAS scale. In: *Clinical Examination of Voice*. Wien: Springer Verlag; 1981.
46. Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *J Voice*. 2004;18:299–304.
47. Webb A, Carding P, Deary I, et al. The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Otorhinolaryngol Head Neck*. 2004;261:429–434.
48. Lee M, Drinnan M, Carding P. The reliability and validity of patient self-rating of their own voice quality. *Clin Otolaryngol*. 2005;30:357–361.
49. Peppard RC, Bless DM, Milenkovic P. Comparison of young adult singers and nonsingers with vocal nodules. *J Voice*. 1988;2:250–260.
50. 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.
51. Hogikyan ND, Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *J Voice*. 1999;13:557–569.
52. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care*. 1992;473–483.
53. Bielamowicz S, Kreiman J, Gerratt BR, et al. Comparison of voice analysis systems for perturbation measurement. *J Speech Lang Hear Res*. 1996;39:126–134.
54. Karnell M, Scherer R, Fischer L. Comparison of acoustic voice perturbation measures among three independent voice laboratories. *J Speech Hear Res*. 1991;34:781–790.
55. Wuyts FL, De Bodt MS, Molenberghs G, et al. The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach. *J Speech Lang Hear Res*. 2000;43:796–809.
56. Colton R, Casper J. *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. Baltimore: Williams & Wilkins; 1996.
57. Stemple J, Glaze L, Klaben B. *Clinical Voice Pathology: Theory and Management*. 3rd ed. San Diego: Singular Publishing Group; 2000.
58. Yumoto E, Sasaki Y, Okamura H. Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. *J Speech Lang Hear Res*. 1984;27:2–6.
59. Bough ID, Heuer RJ, Sataloff RT, et al. Intrasubject variability of objective voice measures. *J Voice*. 1996;10:166–174.
60. Sapienza CM. Aerodynamic influences in voice production. *Curr Opin Otolaryngol Head Neck Surg*. 1998;6:156–160.
61. Smitheran JR, Hixon TJ. A clinical method for estimating laryngeal airway resistance during vowel production. *J Speech Hear Disord*. 1981;46:138–146.
62. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice*. 2nd ed. San Diego, CA: Singular Publishing Group; 2000.
63. Yiu EM, Yuen YM, Whitehill T, et al. Reliability and applicability of aerodynamic measures in dysphonia assessment. *Clin Linguist Phon*. 2004;18:463–478.
64. Lee L, Stemple JC, Kizer M. Consistency of acoustic and aerodynamic measures of voice production over 28 days under various testing conditions. *J Voice*. 1999;13:477–483.
65. Netsell R. Subglottal and intraoral air pressures during the intervocalic contrast of/t/and/d. *Phonetica*. 1969;20:68–73.
66. Netsell R, Lotz W, DuChane A, et al. Vocal tract aerodynamics during syllable productions: normative data and theoretical implications. *J Voice*. 1991;5:1–9.
67. Bless DM, Hirano M, Feder R. Videostroboscopic evaluation of the larynx. *Ear Nose Throat J*. 1987;66:289.
68. Peppard RC, Bless DM. A method for improving measurement reliability in laryngeal videostroboscopy. *J Voice*. 1990;4:280–285.
69. Wendler J. Stroboscopy. *J Voice*. 1992;6:149–154.
70. Poburka BJ. A new stroboscopy rating form. *J Voice*. 1999;13:403–413.

71. Rosen CA. Stroboscopy as a research instrument: development of a perceptual evaluation tool. *Laryngoscope*. 2005;115:423–428.
72. Sulter AM, Schutte HK, Miller DG. Standardized laryngeal videostroboscopic rating: differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency, and age. *J Voice*. 1996;10:175–189.
73. STROBE. University of Bern; 2009.
74. Sayles CL. *The effects of vocal function exercises on aerodynamic parameters for children receiving voice lessons*. Miami University; 2003.
75. Mullen R. The state of the evidence: ASHA develops levels of evidence for communication sciences and disorders. *ASHA Leader*. 2007;12:8–9.
76. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14:365–376.