



# Measurement properties of self-report questionnaires on health-related quality of life and functional health status in dysphonia: a systematic review using the COSMIN taxonomy

Renée Speyer<sup>1,2,3</sup> · Jae-Hyun Kim<sup>4</sup> · Kenji Doma<sup>5</sup> · Yu-Wei Chen<sup>6</sup> · Deborah Denman<sup>3</sup> · Debra Phyland<sup>7</sup> · Lauren Parsons<sup>3</sup> · Reinie Cordier<sup>3</sup>

Accepted: 10 September 2018 / Published online: 6 October 2018  
© Springer Nature Switzerland AG 2018

## Abstract

**Purpose** The current review was conducted to identify all self-report questionnaires on functional health status (FHS) and/or health-related quality-of-life (HR-QoL) in adult populations with dysphonia (voice problems), and to evaluate the psychometric properties of the retrieved questionnaires.

**Methods** A systematic review was performed in the electronic literature databases PubMed and Embase. The psychometric properties of the questionnaires were determined using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) taxonomy and checklist. Responsiveness was outside the scope of this review and as no agreed ‘gold standard’ measures are available in the field of FHS and HR-QoL in dysphonia, criterion validity was not assessed. Only questionnaires developed and published in English were included.

**Results** Forty-eight studies reported on the psychometric properties of 15 identified questionnaires. As many psychometric data were missing or resulted from biased study designs or statistical analyses, only preliminary conclusions can be drawn. Based on the current available psychometric evidence in the literature, the Voice Handicap Index seems to be the most promising questionnaire, followed by the Vocal Performance Questionnaire.

**Conclusions** More research is needed to complete missing data on psychometric properties of existing questionnaires in FHS and/or HR-QoL. Further, when developing new questionnaires, the use of item response theory is preferred above classical testing theory, as well as international consensus-based psychometric definitions and criteria to avoid bias in outcome data on measurement properties.

**Keywords** Dysphonia · Voice · Reliability · Validity · Assessment

## Abbreviations

COSMIN	Consensus-based standards for the selection of health measurement instruments
EASE	Evaluation of the ability to sing easily
FHS	Functional health status
GFI	Glottal Function Index

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11136-018-2001-6>) contains supplementary material, which is available to authorized users.

✉ Renée Speyer  
renee.speyer@iso.edu.no

<sup>1</sup> Department of Special Needs Education, University of Oslo, Oslo, Norway

<sup>2</sup> Department of Otorhinolaryngology and Head and Neck Surgery, Leiden University Medical Center, Leiden, The Netherlands

<sup>3</sup> School of Occupational Therapy, Social Work and Speech Pathology, Curtin University, Perth, Australia

<sup>4</sup> Department of Linguistics, Macquarie University, Sydney, Australia

<sup>5</sup> College of Healthcare Sciences, James Cook University, Townsville, Australia

<sup>6</sup> Faculty of Health Sciences, The University of Sydney, Sydney, NSW, Australia

<sup>7</sup> Department of Surgery, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

HR-QoL	Health-related quality of life
SVHI	Singing Voice Handicap Index
SVHI-10	Singing Voice Handicap Index-10
TSEQ	Transgender Self-Evaluation Questionnaire
TVQ <sup>MIF</sup>	Transsexual Voice Questionnaire—Male to Female
VCQ	Voice Capabilities Questionnaire
VDCQ	Voice Disability Coping Questionnaire
VFI	Vocal Fatigue Index
VHI or VHI-30	Voice Handicap Index
VHI-10	Voice Handicap Index-10
VoiSS	Voice and Symptom Scale
VPQ	Vocal Performance Questionnaire
V-RQOL	Voice-related quality of life
VRS	Voice Rating Scale

## Background

Voice problems or dysphonia can be defined as any deviation in voice quality, pitch or loudness, inappropriate for an individual's age, gender or cultural background. Dysphonia can result from alterations in respiratory, laryngeal or vocal tract mechanisms (organic dysphonia), improper or inefficient use of the vocal mechanism (functional dysphonia) or psychological stressors (psychogenic dysphonia) [1–3]. Voice problems can manifest in different voice modalities (e.g. speaking voice, singing voice and shouting voice) [4]. The burden of dysphonia, its impact on quality of life and work-related effects are increasingly recognised [5].

The prevalence of dysphonia in the general population has been estimated at 0.98% [5]; however, prevalence rates are highly dependent on variables such as gender, age or occupational factors. The lifetime prevalence of a voice disorder may be as high as 29.9% [6], with even higher risks in people for whom using their voice is critical to their vocation, such as teachers [7]. Prevalence data also differ because of variations in the instruments used for measurement. Most researchers and clinicians agree on the fact that voice is a multidimensional phenomenon and follow the guidelines for functional assessment of voice pathology laid out by the Committee on Phoniatics of the European Laryngological Society [8]. In these guidelines, a multidimensional set of minimal basic measurements for all 'common' dysphonias is proposed, involving five different approaches: perception, videostroboscopy, acoustics, aerodynamics and subjective rating by the patient [8]. Still, having reached a consensus on approaches does not imply an agreement on the measures to use in the assessment protocol for voice pathology. Most importantly, the use of a measure in research or clinical practice can only be justified by having robust psychometric

properties: reliability, validity and its discriminative and evaluative ability [9].

Subjective ratings in persons with dysphonia include self-report questionnaires on health-related quality of life (HR-QoL) and functional health status (FHS). HR-QoL is a term that specifically refers to health-related aspects of quality of life, generally considered to reflect the impact of disease and treatment on disability and daily functioning [10]. HR-QoL is the unique personal perception an individual has of his or her health, taking into account social, functional and psychological factors [11]. HR-QoL, however, differs from quality of life as quality of life is a broader construct that encompasses more aspects than just health; that is, quality of life refers to an individuals' perception of their position in life in the context of the culture in which they live and in relation to their goals, expectations, standards and concerns [12, 13]. FHS in turn, refers to the influence of a given disease on particular functional aspects [11]. Function is an umbrella term encompassing all body functions, activities and participation [14]. The health state of an individual at a particular point in time, may be modified by functional states, impairments, perceptions and social opportunities that are influenced by disease, injury treatment and health policy [15]. Still, the distinction between both HR-QoL and FHS concepts can become blurred in measurement. For instance, many self-report questionnaires in dysphonia frequently include items or subscales related to both FHS and HR-QoL. As such, distinction between these concepts is often not possible when using these questionnaires, whereas reduced HR-QoL or impaired FHS may require a different management or intervention.

To select appropriate measures from the available self-report questionnaires, the psychometric properties of each questionnaire must be evaluated and compared. Clinicians and researchers need to carefully select measures with optimal psychometric properties to ensure adequate quality and appropriate interpretation of results. Measures lacking robust psychometric quality cannot guarantee sufficient reliability and validity of retrieved results, thus opposing evidence-based clinical practice and research. The COSMIN group (COnsensus-based Standards for the selection of health Measurement INstruments) established an international consensus-based taxonomy, terminology and definitions of measurement properties for health-related patient-reported outcomes [16]. The framework comprises nine measurement properties subsumed within three domains: reliability, validity and responsiveness. In addition, the COSMIN checklist was developed, providing a standardised and validated tool to rate the methodological quality of studies describing the psychometric properties of self-reported measures in health [17]. The COSMIN framework and checklist have been used in over 560 psychometric reviews (see website: database.cosmin.nl/) and is grounded in contemporary literature, thus

representing the most appropriate methodology to address the psychometric properties of self-report questionnaires in dysphonia.

A simplified checklist to operationalise measurement characteristics of patient-reported outcome measures was developed for reviewers, researchers and clinicians with varied expertise in psychometrics and/or clinimetrics [18]. However, the authors' criteria stand in contrast to the COSMIN checklist which is a complex tool that requires users to have expertise in psychometrics. Furthermore, the simplified checklist shows several methodological shortcomings and received robust critique from the COSMIN group [19]. First, the methodological quality of studies should be distinguished from the effect sizes in trials and separated from the quality of the patient-reported measure itself. This is not the case for the simplified checklist. The results of studies with insufficient methodological quality, may be biased. Therefore, in line with Cochrane methodology, the methodological quality of studies on measurement properties needs to be rated before rating study results; the precise purpose of the COSMIN checklist [19]. In addition, the evaluative criteria of the simplified checklist do not provide sufficient detail for unbiased and systematic rating of the quality of the measure, due to its simplicity [18]. For example, criteria are lacking on what constitutes good content validity, dimensionality or responsiveness. As this checklist was developed for users with limited methodological background, a lack of clarity and standardisation in rating introduces bias upon what constitutes good measurement properties [19].

A psychometric review on voice-related patient-reported outcome measures using the aforementioned simplified checklist was recently published [20]. Due to methodological shortcomings inherent to this checklist, the psychometric properties of self-reported questionnaires in dysphonia remain unclear. No other psychometric reviews in the field of dysphonia have been published.

## Study aim

This systematic review aimed to identify all current self-report questionnaires on FHS and/or HR-QoL in dysphonia for adult populations, and to evaluate the psychometric properties of these questionnaires using the COSMIN framework and checklist.

## Methods

The PRISMA statement [21] and the COSMIN [16, 17] guided the methodology and reporting of this systematic review. This review consists of three consecutive steps: (1) performing a systematic literature search; (2) rating the methodological quality of studies reporting on psychometric properties using

the COSMIN checklist [17]; and (3) rating the quality of each measurement property for all questionnaires using pre-defined criteria [22, 23].

## Eligibility criteria

Self-evaluation questionnaires on FHS and/or HR-QoL in dysphonia, as well as research articles and manuals reporting on the psychometric properties of FHS and/or HR-QoL questionnaires were considered for inclusion in this review. Only questionnaires developed and published in English and research articles and manuals written in English were eligible for inclusion. Questionnaires targeting adults with dysphonia were included, but questionnaires focussing on vocal training were excluded from this review as these measures target a different population. Persons with dysphonia have voice disorders, whereas professional voice users mainly aim at voice optimisation [24]. Single-item questionnaires were excluded as FHS and HR-QoL are complex constructs that cannot be captured by single items only. Furthermore, questionnaires that were not a comprehensive measure were also excluded; for measures to be considered comprehensive, they needed to produce an overall or summative score that was calculated based on the reporting of a number of individual items that collectively comprise the construct of dysphonia.

A minimum of 50% of all items within a questionnaire were required to target the measurement of FHS and/or HR-QoL in dysphonia for the measure to be included. Conference abstracts, reviews, student dissertations and editorials were not considered for inclusion.

## Literature searches and study selection

Systematic literature searches were performed in two different databases: Embase and PubMed. First, databases were searched for self-evaluation questionnaires on FHS and/or HR-QoL in dysphonia (see Supplementary Table 1). Next, additional searches were conducted to identify publications on the psychometric properties of the retrieved questionnaires (see Supplementary Table 2). The final searches were conducted in November 2016; all publications that met the eligibility criteria and were published before this date were included. Two independent reviewers performed the abstract and article selection process. Discrepancies in abstract selection were resolved by consensus between both reviewers. Differences in the final selection of questionnaires or research articles were resolved by group consensus.

## Methodological quality assessment of studies on psychometric properties

To evaluate the methodological quality of the selected studies on psychometric properties, the COSMIN taxonomy of

measurement properties and definitions for health-related patient-reported outcomes was used [16]. The COSMIN framework comprises nine measurement properties: internal consistency, reliability (including test–retest, inter-rater and intra-rater reliability), measurement error, content validity (including face validity), structural validity, hypothesis testing, cross-cultural validity and criterion validity. Table 1 presents definitions for each measurement property used for this review, as guided by the COSMIN statement [16]. Interpretability is not considered to be a psychometric property within the COSMIN framework and was therefore excluded from this review. Responsiveness was outside the scope of this review, and as only original English questionnaires were included, cross-cultural validity was not evaluated. Criterion validity could not be assessed due to the lack of a ‘gold standard’ measure in the field of FHS and HR-QoL in dysphonia.

The COSMIN checklist [17] is a standardised tool and was used to rate the methodological quality of the studies describing the psychometric properties of the included questionnaires. Each measurement property is rated individually, and the checklist for each measurement property contains 5–18 items rated on a four-point scale (poor, fair, excellent, good). The items rate the quality of study design and the robustness of statistical analyses performed in studies on the domains reliability, validity and responsiveness. When using a ‘worst rating counts’ system, the final quality rating for a measurement property is equivalent to the lowest rating given to any of the items contained in the checklist for that property [25]. As this method impedes the detection of subtle differences in methodological quality between studies, a revised scoring procedure was developed [26–28], and this is the method used in this review. Final quality ratings for measurement properties are presented as a percentage using the following formula:

Total score per psychometric property

$$= \frac{(\text{Total score obtained} - \text{Min score possible})}{(\text{Max score possible} - \text{Min score possible})} \times 100\%.$$

The total percentage score is then categorised as poor (0–25%), fair (25.1–50%), good (50.1–75%) or excellent (75.1–100%). Two independent raters with expertise in COSMIN scoring, completed all ratings. To ensure consistency in scoring, a random selection of 40% of all articles retrieved was rated by both raters. The inter-rater reliability was determined by calculating the weighted Kappa between raters.

### Quality of measurement properties

Once the methodological quality of the included studies was determined, the quality of the measurement properties

was evaluated. Research articles that received a poor COSMIN rating were excluded from further analysis. To address the quality of the measurement properties of each questionnaire, psychometric data were retrieved from the selected research studies and were rated according to predefined quality criteria per measurement property [22, 23] (see Supplementary Table 3). Measurement properties could receive a positive, negative, or indeterminate rating. In cases of methodological issues, such as problems in study design or statistical analyses, ratings were classified as indeterminate.

### Overall quality of psychometric properties

Finally, an overall quality score for each measurement property evaluated for each assessment was determined using criteria [23]; these levels of evidence combine the COSMIN ratings for assessing the methodological quality of studies on psychometric properties, and the corresponding quality assessment of psychometric data retrieved from these studies. As a result, an overall quality rating per psychometric property for each questionnaire can be obtained.

## Results

### Systematic literature search

The first systematic literature searches identified self-evaluation questionnaires on FHS and/or HR-QoL related to dysphonia. After deletion of duplicates, a total of 2214 abstracts from Embase (1118 records) and PubMed (1487 records) were identified. Figure 1 presents the flow diagram according to PRISMA [29]. A total of 67 questionnaires were assessed for eligibility, resulting in 15 questionnaires meeting all inclusion criteria. Supplementary table 4 provides a list of the 52 excluded measures and reasons for exclusion.

Additional searches were conducted to retrieve publications on the psychometric properties of the included questionnaires, resulting in a total of 937 abstracts (excluding duplicates): 334 records from Embase and 731 records from PubMed. Data on psychometric properties were retrieved from the literature for all questionnaires. Forty-eight articles reported on at least one or more psychometric properties of any of the 15 questionnaires on FHS and/or HR-QoL in dysphonia. No manuals of questionnaires were located.

### Measures of FHS and HR-QoL in dysphonia

The following 15 questionnaires on FHS and/or HR-QoL were identified: evaluation of the ability to sing easily

**Table 1** Definitions of measurement properties for health-related patient-reported outcomes instruments according to COSMIN [16]

Domain	Measurement property	Aspect of measurement property
Reliability: degree to which the measurement is free from measurement error	Internal consistency: degree of the interrelatedness among the items	
	Reliability: proportion of the total variance in the measurements which is because of “true” differences among patients	
	Measurement error: systematic and random error of a patient’s score that is not attributed to true changes in the construct to be measured	
Validity: degree to which an instrument measures the construct(s) it purports to measure	Content validity: degree to which the content of an instrument is an adequate reflection of the construct to be measured	Face validity: degree to which an instrument indeed looks as though they are an adequate reflection of the construct to be measured
	Construct validity: Degree to which the scores of an instrument are consistent with hypotheses based on the assumption that an instrument validly measures the construct to be measured	Structural validity: degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured Hypotheses testing: idem construct validity Cross-cultural validity: degree to which the performance of the items on a translated or culturally adapted instrument is an adequate reflection of the performance of the items of the original version of the instrument
Responsiveness: ability of an instrument to detect change over time in the construct to be measured	Criterion validity: degree to which the scores of an instrument is an adequate reflection of a ‘gold standard’	
Interpretability <sup>a</sup> : degree to which one can assign qualitative meaning to an instrument’s quantitative scores or change in scores		

<sup>a</sup>Interpretability is not considered a psychometric property

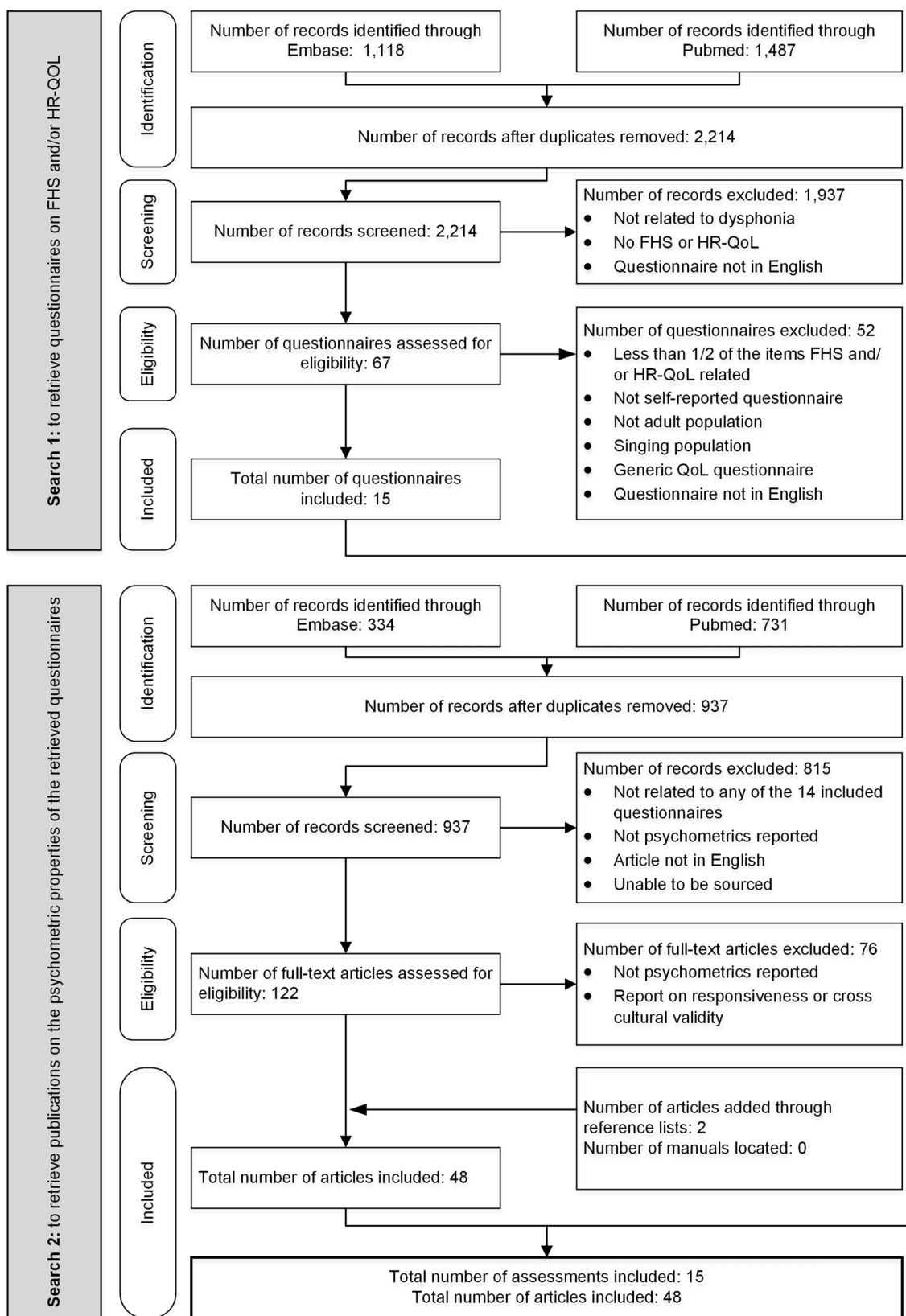


Fig. 1 Flow diagram of the reviewing process according to PRISMA



(EASE) [30], Glottal Function Index (GFI) [31], Singing Voice Handicap Index (SVHI) [32], Singing Voice Handicap Index-10 (SVHI-10) [33], Transgender Self-Evaluation Questionnaire (TSEQ) [34], Transsexual Voice Questionnaire—Male to Female (TVQ<sup>MtF</sup>) [35], Vocal Fatigue Index (VFI) [36], Vocal Performance Questionnaire (VPQ) [37], Voice Capabilities Questionnaire (VCQ) [38], Voice Disability Coping Questionnaire (VDCQ) [39], Voice Handicap Index (VHI or VHI-30) [40], Voice Handicap Index-10 (VHI-10) [41], Voice Rating Scale (VRS) [42], Voice-Related Quality of Life (V-RQOL) [43] and Voice Symptom Scale (VoiSS) [44]. Nine questionnaires combined items on HR-QoL and FHS (EASE, SVHI, SVHI-10, TSEQ, TVQ<sup>MtF</sup>, VHI, VHI-10, V-RQOL and VoiSS), and six questionnaires mainly focussed on FHS (GFI, VFI, VPQ, VCQ and VRS) or HR-QoL (VDCQ). All fifteen questionnaires targeted persons with dysphonia of which three questionnaires aimed at the singing voice (EASE, SVHI and SVHI-10) and two questionnaires at the transgender population (TSEQ and TVQ<sup>MtF</sup>) in particular. Two questionnaires were shorter versions of the original questionnaires: the SVHI-10 and the VHI-10 were shorter versions of the SVHI and VHI, respectively.

Details on the 48 studies on the development and validation of the included questionnaires on FHS and/or HR-QoL in dysphagia are summarised in Supplementary Table 5. Supplementary Table 6 summarises the characteristics of all 15 questionnaires, including names and number of subscales, number of items and response options. Eight questionnaires have no subscales, five questionnaires have three subscales (EASE, TSEQ, VFI, VHI, VoiSS), one questionnaire has two (V-RQOL) subscales and one questionnaire has four (VDCQ) subscales. No cut-off scores are used in any of the included questionnaires (for example, to distinguish between normal voice and dysphonia). All but one questionnaire use a Likert response scale as response option, whereas only the VRS uses visual analogue scales. The total number of items varies between 4 and 36.

### Methodological quality assessment

The COSMIN checklist [17] was used to assess the methodological quality of the 48 included studies. Supplementary Table 7 presents an overview of all COSMIN ratings. Studies that described the psychometric properties of more than one questionnaire were rated multiple times, for each questionnaire separately. Only two studies received poor COSMIN ratings [45, 46] of which one study received a poor rating for one of the analyses (hypothesis testing) [46]. Data resulting from analyses with poor COSMIN ratings were excluded from further analysis, thus leaving 47 studies. All remaining studies were rated as having sufficient methodological quality for further analysis. All studies but four reported on hypothesis

testing. Limited information was retrieved on internal consistency (18 studies), reliability (9 studies, mainly intra-rater reliability), content validity (12 studies) and structural validity (9 studies). No data were identified on measurement error. The inter-rater reliability between both COSMIN raters was very good: weighted Kappa 0.93 (95% CI 0.84–1.00).

### Quality of measurement properties of assessments

Supplementary Table 8 presents the quality of the psychometric properties retrieved from 47 included research articles for all 15 questionnaires based on pre-defined quality criteria [22, 23]. Details on rating criteria are summarised in Supplementary Table 3. The overall, integrated quality score for each measurement property per questionnaire was determined using the criteria or levels of evidence [23], and is presented in Table 2. The overall level of psychometric quality is determined by integrating the methodological quality ratings of the included studies using the COSMIN checklist (Supplementary Table 7) with the quality criteria for measurement properties of the questionnaires [22, 23] (Supplementary Table 8).

None of the measures reported psychometric data on all six measurement properties. In particular, data on measurement error were lacking for all fifteen measures. In total, 42% (38/90) of all six psychometric properties for all fifteen measures were not reported on and 32.7% (17/52) of scores that were reported on were classified as indeterminate ratings. All but two measures (VCQ and VRS) showed positive overall quality scores for at least one psychometric property, whereas four measures (SVHI, SVHI-10, VFI and VoiSS) received a negative overall quality score for a single measurement property. Two measures showed conflicting psychometric data for a single property (VHI-10 and V-RQOL).

### Discussion

The purpose of this systematic review was to identify self-report questionnaires measuring FHS and/or HR-QoL related to dysphonia for adult populations, and to determine the quality of their psychometric properties according to the COSMIN taxonomy.

### Findings on psychometric properties

This review identified 15 questionnaires and 48 studies describing at least one psychometric property of one or more of the included questionnaires. No manuals were retrieved. Of those studies with sufficient methodological

**Table 2** Overall quality score per measurement property per questionnaire

Questionnaire (alphabetical order)	Measurement property: methodological quality per study					
	Internal Consistency	Reliability	Measurement Error	Content Validity	Structural Validity	Hypothesis Testing
Evaluation of the Ability to Sing Easily (EASE) (30)	Strong (Positive result)	NR	NR	Indeterminate	Strong (Positive result)	Indeterminate
Glottal Function Index (GFI) (31)	NR	Indeterminate	NR	NR	NR	Moderate (Positive result)
Singing Voice Handicap Index (SVHI) (32)	Moderate (Negative result)	Indeterminate	NR	Indeterminate	Limited (Positive result)	Limited (Positive result)
Singing Voice Handicap Index-10 (SVHI-10) (33)	Moderate (Positive result)	Indeterminate	NR	Limited (Negative result)	NR	Limited (Positive result)
Transgender Self-Evaluation Questionnaire (TSEQ) (34)	NR	NR	NR	NR	NR	Moderate (Positive result)
Transsexual Voice Questionnaire - Male to Female (TVQ <sup>MtF</sup> ) (35)	Indeterminate	Strong (Positive result)	NR	NR	NR	NR
Vocal Fatigue Index (VFI) (36)	Strong (Positive result)	Indeterminate	NR	Strong (Positive result)	Moderate (Negative result)	Moderate (Positive result)
Vocal Performance Questionnaire (VPQ) (37)	Limited (Positive result)	Strong (Positive result)	NR	NR	Limited (Positive result)	Limited (Positive result)
Voice Capabilities Questionnaire (VCQ) (38)	Indeterminate	NR	NR	NR	NR	NR
Voice Disability Coping Questionnaire (VDCQ) (39)	Indeterminate	NR	NR	Limited (Positive result)	Limited (Positive result)	Indeterminate
Voice Handicap Index (VHI or VHI-30) (40)	Moderate (Positive result)	Moderate (Positive result)	NR	Indeterminate	Moderate (Positive result)	Strong (Positive result)
Voice Handicap Index-10 (VHI-10) (41)	Limited (Positive result)	NR	NR	Strong (Positive result)	Limited (Positive result)	Conflicting
Voice Rating Scale (VRS) (42)	NR	Indeterminate	NR	NR	NR	Indeterminate
Voice-Related Quality of Life (V-RQOL) (43)	Indeterminate	Limited (Positive result)	NR	NR	Moderate (Positive result)	Conflicting
Voice Symptom Scale (VoiSS) (44)	Moderate (Positive result)	Strong (Negative result)	NR	Indeterminate	Strong (Positive result)	Indeterminate

Based on the levels of evidence [23]

quality (47 studies), 12 studies determined psychometric properties of more than one questionnaire, including hypotheses testing describing associations between two of the included questionnaires. The number of psychometric properties per questionnaire addressed in each study was limited. Most studies (31 of 47; 66%) addressed a single psychometric property; however, 10 of the 15 questionnaires had evaluated four or more psychometric properties. Furthermore, 47% (48 of 102) of all quality ratings on psychometric properties retrieved from the 47 studies were classified as indeterminate, which resulted in 33% (17 of 52) of the overall quality scores per psychometric property per questionnaire being classified as indeterminate. Therefore, when describing the psychometric characteristics of FHS and/or HR-QoL questionnaires in dysphonia, many data in the literature are lacking or remain unclear due to methodological or statistical flaws in the identified psychometric studies. As a consequence, the findings from this systematic review indicate an incomplete psychometric overview and the generalisability and interpretation of results remain limited.

For two questionnaires, only data on a single psychometric characteristic were retrieved and for another three questionnaires data were found on two characteristics. For six questionnaires data were reported on for four psychometric characteristics, and for four questionnaires data were reported on for five characteristics. Hypotheses testing was most frequently determined (13 of 15), next internal consistency (12 of 15) and reliability (10 of 15), followed by structural validity (9 out of 15) and content validity (8 of 15). For all but two questionnaires [35, 38], data were retrieved

for at least one aspect of validity (content validity, structural validity or hypotheses testing). No data were identified on measurement error for any of the questionnaires. Responsiveness was out of the scope of this review; cross-cultural and criterion validity were also not determined as only questionnaires developed and published in English were included and no ‘gold standard’ instrument for FHS and/or HR-QoL in dysphonia was identified.

Based on the available psychometric data for the 15 included questionnaires and excluding those questionnaires with negative (SVHI, SVHI-10, VFI, VoiSS) or conflicting ratings (VHI-10, V-RQOL), the VHI seemed to be the most promising questionnaire. The VHI showed strong positive evidence for hypotheses testing, moderate positive evidence for three further properties (internal consistency, reliability, structural validity) and an indeterminate rating for content validity. Next best was the VPQ with strong positive evidence on reliability and limited positive evidence on three other properties (internal validity, structural validity, hypothesis testing). The EASE and VDCQ showed positive evidence on two psychometric properties. For the EASE strong positive evidence was found for internal consistency and structural validity, and indeterminate ratings for content validity and hypothesis testing. The VDCQ showed limited positive ratings for content validity and structural validity, and had indeterminate ratings for internal consistency and hypotheses testing. Three questionnaires received positive ratings for a single property: the TVQ<sup>MtF</sup> (strong positive rating for reliability and indeterminate rating for internal consistency), the GFI (moderate positive rating



for hypothesis testing and indeterminate rating for reliability) and the TSEQ (moderate positive rating for hypothesis rating).

In addition to its psychometric properties, the reasons for selecting a questionnaire may depend on clinical or research purposes. Therefore, population-specific measures may be preferred, such as questionnaires targeting singers (EASE, SVHI or SVHI-10) or male-to-female transsexual women (TSEQ or TVQ<sup>MtF</sup>). The included questionnaires [15] differed in target populations, instrument purposes and measure format, including the number of subscales, items and response options. Despite these differences, all measures had in common that they were self-report questionnaires on FHS and/or HR-QoL in dysphonia in adult populations and, therefore, included in this review. The psychometric evidence on the measurement properties of the questionnaire needs to be considered before a final decision can be made as to which questionnaire to select. Incomplete data on psychometric properties of questionnaires do not necessarily imply poor psychometric quality; however, selection of these questionnaires is not currently supported by robust evidence. This lack of psychometric data on many of the questionnaires on FHS and/or HR-QoL is therefore concerning. For example, if no psychometric data are available on content validity, doubt may arise as to whether the content of the questionnaire adequately reflects the construct being evaluated. This would contradict the use of the questionnaire as content validity is considered as one of the most important measurement properties [47]. Likewise, the use of a questionnaire with negative psychometric evidence cannot be justified based on its psychometric properties.

A recent review on the psychometric properties of voice-related patient-reported measures [20] summarised findings based on a newly developed simplified checklist (present or absent) of evaluative criteria to operationalise measurement characteristics [18]. The COSMIN group criticised the methodological shortcomings of this checklist strongly [19], indicating that the psychometric properties of questionnaires in dysphonia remain unclear based on the review using the simplified checklist [18]. Given that the inclusion and exclusion criteria of this review differed to that of the review using the simplified checklist [20], eight of the 15 included questionnaires in the current review were also reviewed in the previous review [20]: GFI, VFI, VPQ, VDCQ, VHI, VHI-10, V-RQOL and VoiSS. When considering the eight measures that overlap between both reviews, the previous review [20] favours the use of the V-RQOL over the use of the other seven measures when considering their developmental measurement properties and applicability. In contrast, the V-RQOL only achieved limited and moderate positive ratings for two psychometric properties, and indeterminate and conflicting ratings for another two properties, respectively, when using the COSMIN taxonomy. When

evaluating the psychometric properties of the same eight measures using the COSMIN taxonomy, the VHI followed by the VPQ were found to be the most promising measures. Even though the previous review [20] rated the VHI as the equal fourth best measure, the VPQ received a much lower rating (second lowest rating within the selected eight measures when using COSMIN). These findings indicate that not only does the simplified checklist have methodological shortcomings as outlined by the COSMIN group [19], its use leads to different results compared to the COSMIN taxonomy. The terminology, interpretation of identified psychometric data and overall quality ratings for measurement properties using the simplified checklist [20] differ substantially from the psychometric data reported in the current review using the COSMIN taxonomy and checklist. Our findings contra-indicate the use of the simplified checklist [18] for evaluating the psychometric properties of measures for clinical and research purposes.

Remarkably, all but one study in this review used classical testing theory (CTT). Only one study [30] used the more recently developed item response theory (IRT) to determine psychometric properties. Even though the methodologies and interpretation of CTT findings are easier to interpret than those of IRT, the CTT framework has some limitations. In contrast to IRT where the unit of analysis and results are not restricted to the test population, the evaluation of psychometric properties in CTT is specific to the test population. Further, CTT assesses the performance of a measure as a whole, whilst IRT evaluates the reliability of each individual item [48]. The IRT models estimate both item and person parameters within the same model, calculate person-free parameter estimation and item-free trait level estimation, and identify optimal scaling of individual differences based on the evaluation of differential item functioning [49]. Based on the added value of IRT, future studies on the development and validation of measures should consider using IRT instead of CTT.

## Limitations

This review has some limitations; only questionnaires validated in English and psychometric studies published in English were included. Therefore, some psychometric findings on FHS and/or HR-QoL questionnaires in dysphonia may have been excluded. Further, not all authors who published on the psychometric properties of the included questionnaires were contacted. Finally, we did not report on all nine psychometric properties within the COSMIN framework; criterion validity was not included because no agreed gold standard in the field of FHS and/or HR-QoL in dysphonia is available, and responsiveness was out of the scope of our current review. As interpretability is not considered a psychometric property within the COSMIN taxonomy, which

has been confirmed in more recent literature [50], interpretability was also not reported on.

## Conclusions

This systematic review reports on the psychometric properties of 15 self-reported questionnaires for the evaluation of FHS and/or HR-QoL in adults with dysphonia. The COSMIN taxonomy and checklist were used to assess the methodological quality of 48 studies reporting on psychometric characteristics of the included questionnaires. Quality criteria were used to rate the psychometric data on measurement properties for each study [22, 23]. An overall quality score per measurement property per questionnaire was determined by applying the criteria or levels of evidence [23]. Only preliminary conclusions can be drawn as many psychometric data proved missing or indeterminate for all questionnaires included. Based on current available psychometric data from the literature, the VHI seems to be the most promising questionnaire, followed by the VPQ. More research is needed to evaluate the quality of the psychometric properties of existing questionnaires that has not been tested to date, and augment evaluations of questionnaires using both IRT modeling and international consensus-based psychometric quality criteria and terminology, such as the COSMIN framework.

**Acknowledgements** The authors express their gratitude to Jessica Powell, Grace Martin and Melinda Licciardello for providing support during the start-up of this review.

## Compliance with ethical standards

**Conflict of interest** DP authored one of the instruments included in this review (EASE); however, DP was not involved in COSMIN scorings or quality ratings of measurement properties of questionnaires; raters were blinded to the involvement of DP in the paper. All other authors declare that they have no competing interests.

**Research involving human and animal participants** This article does not contain any studies with human participants or animals performed by any of the authors.

## References

- Colton, R. H., Casper, J. K., & Leonard, R. (2011). *Understanding voice problems: A physiological perspective for diagnosis and treatment*. Baltimore: Lippincott Williams & Wilkins.
- Speyer, R. (2008). Effects of voice therapy: A systematic review. *Journal of Voice*, 22(5), 565–580.
- Verdolini, K., & Ramig, L. O. (2001). Review: Occupational risks for voice problems. *Logopedics Phoniatics Vocology*, 26(1), 37–46.
- Hacki, T. (1996). Comparative speaking, shouting and singing voice range profile measurement: Physiological and pathological aspects. *Logopedics Phoniatics Vocology*, 21, 123–129.
- Cohen, S. M., Kim, J., Roy, N., Asche, C., & Courey, M. (2012). Prevalence and causes of dysphonia in a large treatment-seeking population. *The Laryngoscope*, 122, 343–348.
- Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M. (2005). Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *The Laryngoscope*, 115, 1988–1995.
- Roy, N., Merrill, R. M., Thibeault, S., Gray, S. D., & Smith, E. M. (2004). Voice disorders in teachers and the general population: Effects on work performance, attendance, and future career choices. *Journal of Speech, Language and Hearing Research*, 47, 542–551.
- Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., et al. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques: Guideline elaborated by the Committee on Phoniatics of the European Laryngological Society (ELS). *European Archives of Oto-Rhino-Laryngology*, 258, 77–82.
- Speyer, R. (2013). Oropharyngeal dysphagia: Screening and assessment. In: K. W. Altman (Ed.), *Dysphagia otolaryngologic clinics of North America* (Vol. 46, pp. 989–1008). Philadelphia: The clinics: Internal Medicine: Elsevier Health Sciences.
- Kaplan, R. M. (1985). *Quality of life measures: Measurement strategies in health psychology*. New York: Wiley.
- Ferrans, C. E., Zerwic, J. J., Wilbur, J. E., & Larson, J. L. (2005). Conceptual model of health-related quality of life. *Journal of Nursing Scholarship*, 37(4), 336–342.
- WHOQOL. The World Health Organization Quality of Life Assessment (WHOQOL). (1995). Position paper from the World Health Organization. *Social Science & Medicine*, 41(10), 1403–1409.
- Guyatt, G. H., & Cook, D. J. (1994). Health status, quality of life, and the individual. *Journal of the American Medical Association*, 272(8), 630–631.
- WHO. (2001). *International classification of functioning, disability and health* (2nd ed.). Geneva: WHO.
- Gold, M. R., Segel, J. E., Russel, L. B., & Weinstein, M. C. (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737–745.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, 19(4), 539–549.
- Francis, D. O., McPheeters, M. L., Noud, M., Penson, D. F., & Feurer, I. D. (2016). Checklist to operationalize measurement characteristics of patient-reported outcome measures. *Systematic Reviews*, 5(1), 129.
- Terwee, C. B., De Vet, H. C., Prinsen, C. A. C., & Mokkink, L. B. (2016). Comment on “Checklist to operationalize measurement characteristics of patient-reported outcome measures” VU University Medical Center, Department of Epidemiology and Biostatistics, Amsterdam, The Netherlands. <http://www.cosmin.nl/commentary-on-francis-do-syst-revs-.html>.
- Francis, D. O., Daniero, J. J., Hovis, K. L., Sathe, N., Jacobson, B., Penson, D. F., et al. (2017). Voice-related patient-reported outcome measures: A systematic review of instrument

- development and validation. *Journal of Speech, Language & Hearing Research*, 60, 62–88.
21. Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Annals of Internal Medicine*, 151(4), W1–W30.
  22. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34–42.
  23. Schellingerhout, J. M., Verhagen, A. P., Heymans, M. W., Koes, B. W., Henrica, C., & Terwee, C. B. (2012). Measurement properties of disease-specific questionnaires in patients with neck pain: A systematic review. *Quality of Life Research*, 21(4), 659–670.
  24. Timmermans, B., De Bodt, M. S., Wuyts, F. L., & van de Heyning, P. H. (2004). Training outcome in future professional voice users after 18 months of voice training. *Folia Phoniatrica et Logopaedica*, 56, 120–129.
  25. Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, 21(4), 651–657.
  26. Speyer, R., Cordier, R., Kertscher, B., & Heijnen, B. J. (2014). Psychometric properties of questionnaires on functional health status in oropharyngeal dysphagia: A systematic literature review. *BioMed Research International*. <https://doi.org/10.1155/2014/458678>.
  27. Cordier, R., Speyer, R., Chen, Y. W., Wilkes-Gillan, S., Brown, T., Bourke-Taylor, H., et al. (2015). Evaluating the psychometric quality of social skills measures: A systematic review. *PLoS ONE*, 10(7), e0132299.
  28. Cordier, R., Chen, Y.-W., Speyer, R., Totino, R., Doma, K., Leicht, A., et al. (2016). Child-report measures of occupational performance: A systematic review. *PLoS ONE*, 11(1), e0147751.
  29. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269.
  30. Phyland, D. J., Pallant, J. F., Benninger, M. S., Thibeault, S. L., Greenwood, K. M., Smith, J. A., et al. (2013). Development and preliminary validation of the EASE: A tool to measure perceived singing voice function. *Journal of Voice*, 27(4), 454–462.
  31. Bach, K. K., Belafsky, P. C., Wasylik, K., Postma, G. N., & Koufman, J. A. (2005). Validity and reliability of the glottal function index. *Archives of Otolaryngology–Head & Neck Surgery*, 131(11), 961–964.
  32. Cohen, S. M., Jacobson, B. H., Garrett, C. G., Noordzij, J. P., Stewart, M. G., Attia, A., et al. (2007). Creation and validation of the Singing Voice Handicap Index. *Annals of Otolaryngology, Rhinology & Laryngology*, 116(6), 402–406.
  33. Cohen, S. M., Statham, M., Rosen, C. A., & Zullo, T. (2009). Development and validation of the singing voice handicap-10. *The Laryngoscope*, 119(9), 1864–1869.
  34. Davis, S. (2006). Transgender Self-Evaluation Questionnaire. In H. Adler & M. Mordaunt (Eds.), *Voice and communication therapy for the transgender/transsexual client: A comprehensive clinical guide* (pp. 485–487). San Diego: Plural Publishing.
  35. Dacakis, G., Davies, S., Oates, J. M., Douglas, J. M., & Johnston, J. R. (2013). Development and preliminary evaluation of the transsexual voice questionnaire for male-to-female transsexuals. *Journal of Voice*, 27(3), 312–320.
  36. Nanjundeswaran, C., Jacobson, B. H., Gartner-Schmidt, J., & Verdolini Abbott, K. (2015). Vocal Fatigue Index (VFI): Development and validation. *Journal of Voice*, 29(4), 433–440.
  37. Carding, P. N., & Horsley, I. A. (1992). An evaluation study of voice therapy in non-organic dysphonia. *International Journal of Language & Communication Disorders*, 27(2), 137–158.
  38. Buckley, K. L., O'Halloran, P. D., & Oates, J. M. (2015). Occupational vocal health of elite sports coaches: An exploratory pilot study of football coaches. *Journal of Voice*, 29(4), 476–483.
  39. Epstein, R., Hirani, S. P., Stygall, J., & Newman, S. P. (2009). How do individuals cope with voice disorders? Introducing the Voice Disability Coping Questionnaire. *Journal of Voice*, 23(2), 209–217.
  40. Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., et al. (1997). The Voice Handicap Index (VHI) development and validation. *American Journal of Speech-Language Pathology*, 6(3), 66–70.
  41. Rosen, C. A., Lee, A. S., Osborne, J., Zullo, T., & Murry, T. (2004). Development and Validation of the Voice Handicap Index-10. *The Laryngoscope*, 114(9), 1549–1556.
  42. Wingate, J. M., Brown, W. S., Shrivastav, R., Davenport, P., & Sapienza, C. M. (2007). Treatment outcomes for professional voice users. *Journal of Voice*, 21(4), 433–449.
  43. Hogikyan, N. D., & Sethuraman, G. (1999). Validation of an instrument to measure voice-related quality of life (V-RQOL). *Journal of Voice*, 13(4), 557–569.
  44. Deary, I. J., Wilson, J. A., Carding, P. N., & MacKenzie, K. (2003). VoiSS: A patient-derived Voice Symptom Scale. *Journal of Psychosomatic Research*, 54(5), 483–489.
  45. Arffa, R. E., Krishna, P., Gartner-Schmidt, J., & Rosen, C. A. (2012). Normative values for the Voice Handicap Index-10. *Journal of Voice*, 26(4), 462–465.
  46. Wilson, J. A., Webb, A., Carding, P. N., Steen, I. N., MacKenzie, K., & Deary, I. J. (2004). The Voice Symptom Scale (VoiSS) and the Vocal Handicap Index (VHI): A comparison of structure and content. *Clinical Otolaryngology and Allied Sciences*, 29(2), 169–174.
  47. Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., et al. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*, 10, 22.
  48. Linacre, J. M. (2016). *Rasch-model computer programs: Program manual 3.92.0*. Chicago: Mesa-Press.
  49. Reise, S., & Henson, J. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment*, 81(2), 93–103.
  50. Mokkink, L. B., De Vet, H. C., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., et al. (2018). COSMIN risk of bias checklist for systematic reviews of patient reported outcome measures. *Quality of Life Research*, 27, 1171–1179.
  51. Brutten, G., & Vanryckeghem, M. (2003). *Behavior assessment battery: A multi-dimensional and evidence-based approach to diagnostic and therapeutic decision making for children who stutter*. Leuven: Belgium & Acco Publishers.
  52. Borg, G. (1990). Psychological scaling with applications in physical work and the perception of exertion. *Scandinavian Journal of Work, Environment & Health*, 16(Suppl.1), 55–58.
  53. Ford Baldner, E., Doll, E., & van Mersbergen, M. R. (2015). A review of measures of vocal effort with a preliminary study on the establishment of a vocal effort measure. *Journal of Voice*, 29(5), 530–541.
  54. Ricci-Maccarini, A., De Maio, V., Murry, T., & Schindler, A. (2013). Development and validation of the Children's Voice Handicap Index-10 (CVHI-10). *Journal of Voice*, 27(2), 258.e23–258.e28.

55. Fussi, F. (2005). La valutazione del canto. In F. Fussi (Ed.), *La voce del cantante* (pp. 33–68). Torino: Omega.
56. Baylor, C., Yorkston, K. M., Eadie, T. L., Jiseon, K., Hyewon, C., & Amtmann, D. (2013). The communicative participation item bank (CPIB): Item bank calibration and development of a disorder-generic short form. *Journal of Speech, Language & Hearing Research, 56*(4), 1190–1208.
57. Giannini, S. P., Latorre Mdo, R., & Ferreira, L. P. (2016). Condition of vocal production-teacher questionnaire: Comparison of responses on likert scale and visual analog scale. *Codas, 28*(1), 53–58.
58. Hu, A., Isetti, D., Hillel, A. D., Waugh, P., Comstock, B., & Meyer, T. K. (2013). Disease-specific self-efficacy in spasmodic dysphonia patients. *Otolaryngology–Head and Neck Surgery, 148*(3), 450–455.
59. Nemr, K., Simões-Zenari, M., Duarte, J. M. D. T., Lobrigate, K. E., & Bagatini, F. A. (2016). Dysphonia risk screening protocol. *Clinics, 71*(3), 114–127.
60. Wuyts, F. L., De Bodt, M., Molenberghs, G., Remacle, M., Heylen, L., Millet, B., et al. (2000). The Dysphonia Severity Index: An objective measure of vocal quality based on a multiparameter approach. *Journal of Speech Language and Hearing Research, 43*(3), 796–809.
61. Robinson, K., Gatehouse, S., & Browning, G. G. (1996). Measuring patient benefit from otorhinolaryngological surgery and therapy. *The Annals of Otolaryngology, Rhinology & Laryngology, 105*(6), 415–422.
62. Kubba, H., Swan, I. R. C., & Gatehouse, S. (2004). The Glasgow children's benefit inventory: A new instrument for assessing health-related benefit after an intervention. *Annals of Otolaryngology, Rhinology & Laryngology, 113*(12), 980–986.
63. Goldman, S. L., Hargrave, J., Hillman, R. E., Holmberg, E., Gress, C. (1996). Stress anxiety, somatic complaints, and voice use in women with vocal nodules: Preliminary findings. *American Journal of Speech-Language Pathology, 5*(1), 44–54.
64. Karnell, M. P., Melton, S. D., Childes, J. M., Coleman, T. C., Dailey, S. A., & Hoffman, H. T. (2007). Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *Journal of Voice, 21*(5), 576–590.
65. Dehqan, A., Yadegari, F., Asgari, A., Scherer, R. C., & Dabirmoghdam, P. (2017). Development and validation of an Iranian voice quality of life profile (IVQLP) based on a classic and Rasch Rating Scale Model (RSM). *Journal of Voice, 31*(1), 113–119.
66. Baylor, C., Yorkston, K., Eadie, T., Miller, R., & Amtmann, D. (2008). Levels of speech usage: A self-report scale for describing how people use speech. *Journal of Medical Speech-Language Pathology, 16*(4), 191–198.
67. Llewellyn-Thomas, H. A., Sutherland, H. J., Hogg, S. A., Ciampi, A., Harwood, A. R., Keane, T. J., et al. (1984). Linear analogue self-assessment of voice quality in laryngeal cancer. *Journal of Chronic Diseases, 37*(12), 917–924.
68. Hartelius, L., Elmberg, M., Holm, R., Löfberg, A. S., & Nikolaidis, S. (2008). Living with dysarthria: Evaluation of a self-report questionnaire. *Folia Phoniatrica et Logopaedica, 60*(1), 11–19.
69. Costa, F. P., Diaferia, G., & Behlau, M. (2016). Communicative aspects and coping strategies in patients with Parkinson's disease. *Codas, 28*(1), 46–52.
70. Priebe, S., Huxley, P., Knight, S., & Evans, S. (1999). Application and Results of the Manchester Short Assessment of Quality of Life (MANSA). *International Journal of Social Psychiatry, 45*(1), 7–12.
71. Moreti, F., Rocha, C., Borrego, M. C. M., & Behlau, M. (2011). Desvantagem vocal no canto: Análise do protocolo Índice de Desvantagem para o Canto Moderno—IDCM. *Revista da Sociedade Brasileira de Fonoaudiologia, 16*(2), 146–151.
72. Koufman, J. A., & Isaacson, G. (1991). The spectrum of vocal dysfunction. *Otolaryngologic Clinics of North America, 24*(5), 985–988.
73. Zur, K. B., Cotton, S., Kelchner, L., Baker, S., Weinrich, B., & Lee, L. (2007). Pediatric Voice Handicap Index (pVHI): A new tool for evaluating pediatric dysphonia. *International Journal of Pediatric Otorhinolaryngology, 71*(1), 77–82.
74. Hartnick, C. J. (2002). Validation of a pediatric voice quality-of-life instrument: The pediatric voice outcome survey. *Archives of Otolaryngology–Head & Neck Surgery, 128*(8), 919–922.
75. Boseley, M. E., Cunningham, M. J., Volk, M. S., & Hartnick, C. J. (2006). Validation of the pediatric voice-related quality-of-life survey. *Archives of Otolaryngology–Head & Neck Surgery, 132*(7), 717–720.
76. Smith, E., Verdolini, K., Gray, S., Nichols, S., Lemke, J., Barkmeier, J., et al. (1996). Effect of voice disorders on quality of life. *Journal of Medical Speech-Language Pathology, 4*, 223–244.
77. De Araujo Pernambuco, L., Espelt, A., Virgilio Magalhaes Junior, H., Veiga Andersen Cavalcanti, R., & Costa de Lima, K. (2016). Screening for voice disorders in older adults—Part I: Validity evidence based on test content and response processes. *Journal of Voice, 30*(2), 246.e9–246.e17.
78. Vitali, C., Borghi, E., Napoletano, A., Polini, F., Caronni, M., Ammenti, P., et al. (2010). Oropharyngolaryngeal disorders in scleroderma: Development and validation of the SLS scale. *Dysphagia, 25*(2), 127–138.
79. Harris, T., Collins, S., & Clarke, D. D. (1986). An interdisciplinary voice clinic. In M. Fawcus (Ed.), *Voice disorders and their management* (pp. 283–292). London: Croom Helm.
80. Sheehan, D. V. (1983). *The anxiety disease*. New York: Scribner.
81. Garcia, G. M., Lopez, J. M., & Lara, E. Y. M. (2017). Voice habits and behaviours: Voice care among Flamenco singers. *Journal of Voice, 31*(2), 246.e11–246.e19.
82. Jones, S. M., Carding, P. N., & Drinnan, M. J. (2006). Exploring the relationship between severity of dysphonia and voice-related quality of life. *Clinical Otolaryngology, 31*(5), 411–417.
83. Epstein, R., Stygall, J., & Newman, S. (1997). The short-term impact of Botox injections on speech disability in adductor spasmodic dysphonia. *Disability and Rehabilitation, 19*(1), 20–25.
84. Isetti, D., & Meyer, T. (2014). Workplace productivity and voice disorders: A cognitive interviewing study on presenteeism in individuals with spasmodic dysphonia. *Journal of Voice, 28*(6), 700–710.
85. Hill, D. S., Akhtar, S., Corroll, A., & Croft, C. B. (2000). Quality of life issues in recurrent respiratory papillomatosis. *Clinical Otolaryngology & Allied Sciences, 25*(2), 153–160.
86. Kirsh, E. R., van Leer, E., Phero, H. J., Xie, C., & Khosla, S. (2013). Factors associated with singers' perceptions of choral singing well-being. *Journal of Voice, 27*(6), 786.e25–786.e32.
87. Nam, I., Bae, J., Shim, M., Hwang, Y., Kim, M., & Sun, D. (2012). The importance of preoperative laryngeal examination before thyroidectomy and the usefulness of a voice questionnaire in screening. *World Journal of Surgery, 36*(2), 303–309.
88. Stewart, C. F., Allen, E. L., Tureen, P., Diamond, B. E., Blitzer, A., & Brin, M. F. (1997). Adductor spasmodic dysphonia: Standard evaluation of symptoms and severity. *Journal of Voice, 11*(1), 95–103.
89. Teixeira, L. C., Rodrigues, A. L. V., Silva, Á. F. G. D., Azevedo, R., Gama, A. C. C., & Behlau, M. (2013). The use of the URICA-VOICE questionnaire to identify the stages of adherence to voice treatment. *Codas, 25*(1), 8–15.
90. Speyer, R., Wieneke, G. H., & Dejonckere, P. H. (2004). Self-assessment of voice therapy for chronic dysphonia. *Clinical Otolaryngology, 29*(1), 66–74.



91. Paolillo, N. P., & Pantaleo, G. (2015). Development and validation of the Voice Fatigue Handicap Questionnaire (VFHQ): Clinical, psychometric, and psychosocial facets. *Journal of Voice*, 29(1), 91–100.
92. Mathieson, L., Hirani, S. P., Epstein, R., Baken, R. J., Wood, G., & Rubin, J. S. (2009). Laryngeal manual therapy: A preliminary study to examine its treatment effects in the management of muscle tension dysphonia. *Journal of Voice*, 23(3), 353–366.
93. Ma, E. P., & Yiu, E. M. (2001). Voice activity and participation profile: Assessing the impact of voice disorders on daily activities. *Journal of Speech, Language & Hearing Research*, 44(3), 511–524.
94. Konnai, R. M., Jayaram, M., & Scherer, R. C. (2010). Development and validation of a voice disorder outcome profile for an Indian population. *Journal of Voice*, 24(2), 206–220.
95. Gugatschka, M., Rechenmacher, J., Chibidziura, J., & Friedrich, G. (2007). Vergleichbarkeit und Urechnung von Stimmstörungsindex (SSI) und Voice Handicap Index (VHI). *Laryngo-Rhino-Otologie*, 86(11), 785.
96. Li, H., Huang, Z., Hu, R., Zhang, L., & Xu, W. (2012). Study on the simplified Chinese version of the voice handicap index. *Journal of Voice*, 26(3), 365–371.
97. Zraick, R. I., Risner, B. Y., Smith-Olinde, L., Gregg, B. A., Johnson, F. L., & McWeeny, E. K. (2007). Patient versus partner perception of voice handicap. *Journal of Voice*, 21(4), 485–494.
98. Lyberg-Åhlander, V., Rydell, R., Eriksson, J., & Schälén, L. (2010). Throat related symptoms and voice: Development of an instrument for self assessment of throat-problems. *Ear, Nose and Throat Disorders*, 10(1), 5.
99. El-Banna, M., & Ras, Y. A. (2010). Formulation of a voice problem self assessment scale (VPSS) as a patient based tool for Egyptian dysphonic population. *Alexandria Journal of Medicine*, 46(2), 119–126.
100. Gillespie, A. I., & Abbott, K. V. (2011). The influence of clinical terminology on self-efficacy for voice. *Logopedics Phoniatrics Vocology*, 36(3), 91–99.
101. Gliklich, R. E., Glovsky, R. M., & Montgomery, W. W. (1999). Validation of a voice outcome survey for unilateral vocal cord paralysis. *Otolaryngology–Head and Neck Surgery*, 120(2), 153–158.
102. Reilly, M. C., Zbrozek, A. S., & Dukes, E. M. (1993). The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics*, 4(5), 353–365.
103. Rosow, D. E., Szczupak, M., Saint-Victor, S., Gerhard, J. D., Dupont, C., & Lo, K. (2016). The economic impact of vocal attrition in public school teachers in Miami-Dade County. *Laryngoscope*, 126(3), 665–671.
104. Phyland, D. J., Pallant, J. F., Thibeault, S. L., Benninger, M. S., Vallance, N., & Smith, J. A. (2014). Measuring vocal function in professional music theater singers: Construct validation of the evaluation of the ability to sing easily (EASE). *Folia Phoniatrica et Logopaedica*, 66, 100–108.
105. Buckmire, R. A., Bryson, P. C., & Patel, M. R. (2011). Type I gore-tex laryngoplasty for glottic incompetence in mobile vocal folds. *Journal of Voice*, 25(3), 288–292.
106. Castelblanco, L., Habib, M., Stein, D. J., de Quadros, A., Cohen, S. M., & Noordzij, J. P. (2014). Singing voice handicap and videostrobolaryngoscopy in healthy professional singers. *Journal of Voice*, 28(5), 608–613.
107. Hancock, A. B., Krissinger, J., & Owen, K. (2011). Voice perceptions and quality of life of transgender people. *Journal Voice*, 25(5), 553–558.
108. Hancock, A. B. (2017). An ICF perspective on voice-related quality of life of American transgender women. *Journal Voice*, 31(1), 115.e1–115.e8.
109. Deary, I. J., Webb, A., Mackenzie, K., Wilson, J. A., & Carding, P. N. (2004). Short, self-report voice symptom scales: Psychometric characteristics of the Voice Handicap Index-10 and the vocal performance questionnaire. *Otolaryngology-Head and Neck Surgery*, 131(3), 232–235.
110. Webb, A. L., Carding, P. N., Deary, I. J., MacKenzie, K., Steen, I. N., & Wilson, J. A. (2007). Optimising outcome assessment of voice interventions, I: Reliability and validity of three self-reported scales. *Journal of Laryngology and Otology*, 121(8), 763–767.
111. Awan, S. N., Roy, N., & Cohen, S. M. (2014). Exploring the relationship between spectral and cepstral measures of voice and the Voice Handicap Index (VHI). *Journal of Voice*, 28(4), 430–439.
112. Elam, J. C., Ishman, S. L., Dunbar, K. B., Clarke, J. O., & Gourin, C. G. (2010). The relationship between depressive symptoms and Voice Handicap Index scores in laryngopharyngeal reflux. *Laryngoscope*, 120(9), 1900–1903.
113. Fulljames, N., & Harris, S. (2006). Voice outcome measures: Correlations with patients' assessment of their condition and the effectiveness of voice therapy. *Logoped Phoniatr Vocol*, 31(1), 23–35.
114. Kazi, R., De Cordova, J., Singh, A., Venkitaraman, R., Nutting, C. M., Clarke, P., et al. (2007). Voice-related quality of life in laryngectomees: Assessment using the VHI and V-RQOL symptom scales. *Journal of Voice*, 21(6), 728–734.
115. Portone, C. R., Hapner, E. R., McGregor, L., Otto, K., & Johns, M. M. 3rd (2007). Correlation of the Voice Handicap Index (VHI) and the voice-related quality of life measure (V-RQOL). *Journal of Voice*, 21(6), 723–727.
116. Stomeo, F., Tosin, E., Morolli, F., Bianchini, C., Ciorba, A., Pastore, A., et al. (2013). Comparison of subjective and objective tools in transoral laser cordectomy for early glottic cancer: Importance of voice handicap index. *International Journal of Immunopathology and Pharmacology*, 26(2), 445–451.
117. Wheeler, K. M., Collins, S. P., & Sapienza, C. M. (2006). The relationship between VHI scores and specific acoustic measures of mildly disordered voice production. *Journal of Voice*, 20(2), 308–317.
118. Childs, L. F., Bielinski, C., Toles, L., Hamilton, A., Deane, J., & Mau, T. (2015). Relationship between patient-perceived vocal handicap and clinician-rated level of vocal dysfunction. *Laryngoscope*, 125(1), 180–185.
119. Davis, K. M., Sandage, M. J., Plexico, L., & Pascoe, D. D. (2016). The perception of benefit of vocalization on sport performance when producing maximum effort. *Journal of Voice*, 30(5), 639.e11–639.e16.
120. Eadie, T. L., Lamvik, K., Baylor, C. R., Yorkston, K. M., Kim, J., & Amtmann, D. (2014). Communicative participation and quality of life in head and neck cancer. *Annals of Otolaryngology and Rhinology*, 123(4), 257–264.
121. Gillespie, A. I., Gooding, W., Rosen, C., & Gartner-Schmidt, J. (2014). Correlation of VHI-10 to voice laboratory measurements across five common voice disorders. *Journal Voice*, 28(4), 440–448.
122. Hu, A., Hillel, A., & Meyer, T. (2016). Factors associated with patient-perceived hoarseness in spasmodic dysphonia patients. *Journal of Voice*, 30(6), 769.e23–769.e26.
123. Kupfer, R. A., Cadalli Tatar, E., Barry, J. O., Allen, C. T., & Merati, A. L. (2016). Anatomic Derkay score is associated with voice handicap in laryngeal papillomatosis in adults. *Otolaryngology-Head and Neck Surgery*, 154(4), 689–692.
124. Nichols, B., Bock, J. M., & Blumin, J. H. (2015). Dysphonia in nursing home and assisted living residents: Prevalence and association with frailty. *Journal of Voice*, 29(1), 79–82.
125. Romak, J. J., Orbelo, D. M., Maragos, N. E., & Ekbom, D. C. (2014). Correlation of the Voice Handicap Index-10 (VHI-10)



- and voice-related quality of life (V-RQOL) in patients with dysphonia. *Journal of Voice*, 28(2), 237–240.
126. Stachler, R. J., Schultz, L. R., Nerenz, D., & Yaremchuk, K. L. (2014). PROMIS evaluation for head and neck cancer patients: A comprehensive quality-of-life outcomes assessment tool. *Laryngoscope*, 124(6), 1368–1376.
127. Willis, J., Michael, D. D., Boyer, H., & Misono, S. (2015). Prevalence and severity of dysphonia in patients with cystic fibrosis: A pilot study. *Otolaryngology-Head and Neck Surgery*, 153(1), 88–93.
128. Bornbaum, C. C., Day, A. M., & Doyle, P. C. (2014). Examining the construct validity of the V-RQOL in speakers who use alaryngeal voice. *American Journal of Speech-Language Pathology*, 23(2), 196–202.
129. Kupfer, R. A., Hogikyan, E. M., & Hogikyan, N. D. (2014). Establishment of a normative database for the voice-related quality of life (V-RQOL) measure. *Journal of Voice*, 28(4), 449–451.
130. Murry, T., Medrado, R., Hogikyan, N. D., & Aviv, J. E. (2004). The relationship between ratings of voice quality and quality of life measures. *Journal of Voice*, 18(2), 183–192.
131. Tanner, K., Pierce, J. L., Merrill, R. M., Miller, K. L., Kendall, K. A., & Roy, N. (2015). The quality of life burden associated with voice disorders in sjogren's syndrome. *Annals of Otolaryngology & Laryngology*, 124(9), 721–727.
132. Montgomery, J., Hendry, J., Wilson, J. A., Deary, I. J., & MacKenzie, K. (2016). Pragmatic detection of anxiety and depression in a prospective cohort of voice outpatient clinic attenders. *Clinical Otolaryngology*, 41(1), 2–7.