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[1,2,3], Jae-Hyun Kim[4], Kenji Doma[5], Yu-Wei Chen[6], Deborah Denman[3], Debra Phyland[7], Lauren Parsons[3], Reinie Cordier[3]

[1]Department of Special Needs Education, University of Oslo, Oslo, Norway

[2]Department of Otorhinolaryngology and Head and Neck Surgery, Leiden University Medical Center, Leiden, the Netherlands

[3]School of Occupational Therapy, Social Work and Speech Pathology, Curtin University, Perth, Western Australia

[4]Department of Linguistics, Macquarie University, Sydney, Australia

[5]College of Healthcare Sciences, James Cook University, Townsville, Australia

[6]Faculty of Health Sciences, The University of Sydney, Sydney, NSW, Australia

[7]Department of Surgery, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

Renée Speyer (Corresponding author): renee.speyer@iso.edu.no

Jae-Hyun Kim: jae-hyun.kim@mq.edu.au

Kenji Doma: kenji.doma@jcu.edu.au

Yu-Wei Chen: yu-wei.chen@sydney.edu.au

Deborah Denman: deborah.denman@postgrad.curtin.edu.au

Debra Phyland: debra.phyland@monash.edu

Lauren Parsons: lauren.parsons@curtin.edu.au

Reinie Cordier: reinie.cordier@curtin.edu.au

Abstract

Background: Dysphonia or voice problems is considered a multidimensional phenomenon; self-report rating is considered an essential component of the assessment of dysphonia. 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, 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-seven 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 (VHI) seems the most promising questionnaire, followed by the Vocal Performance Questionnaire (VPQ).

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 (IRT) is preferred above Classical Testing Theory (CTT), as well as the use of international consensus-based psychometric definitions and criteria to avoid bias in outcome data on measurement properties.

Key words: Dysphonia, voice, reliability, validity, assessment

Background

The burden of dysphonia, its impact on quality of life, and work-related effects are increasingly recognised [1]. The prevalence of dysphonia in the general population has been estimated at 0.98% [1]; 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% [2], with even higher risks in people for whom using their voice is critical to their vocation, such as teachers [3].

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 Phoniatrics of the European Laryngological Society [4]. 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. 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 [5].

Subject 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 described as the unique personal perception an individual has of his or her health, taking into account social, functional, and psychological issues. FHS on the other hand, refers to the influence of a given disease on particular functional aspects [6]. The distinction between both concepts can become blurred, especially since self-report questionnaires in dysphonia frequently include items or subscales related to both FHS and HR-QoL.

To select appropriate measures from the available self-report questionnaires, the psychometric properties of each questionnaire must be evaluated and compared. 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 [7]. 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 [8]. 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.

Francis, McPheeters, Noud, Penson, and Feurer [9] developed a new evaluative checklist to operationalise measurement characteristics of patient-reported outcome measures. Developed for reviewers, researchers and clinicians with varied expertise in psychometrics and/or clinimetrics, Francis and colleagues' criteria stand in contrast to the COSMIN checklist which is a complex tool that requires users to have expertise in psychometrics. However, the simplified checklist presented by Francis and colleagues shows several methodological shortcomings and was robustly critiqued by the COSMIN group [10]. 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 checklist presented by Francis and colleagues. 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 [10]. In addition, the evaluative criteria by Francis and colleagues do not provide sufficient detail for unbiased and systematic rating of the quality of the measure, due to its simplicity. 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 [10].

Recently, Francis, Daniero, Hovis, Sathe, Jacobson, Penson, Feurer and McPheeters [11] published a psychometric review on voice-related patient-reported outcome measures using the aforementioned simplified checklist. Due to methodological shortcomings inherent to this checklist, the psychometric properties of self-reported questionnaires in dysphonia remain unclear. Apart from the review by Francis and colleagues, 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. This study focused on validity and reliability; responsiveness was considered outside the scope of this review and due to the lack of 'gold standard' measures in the field of FHS and HR-QoL in dysphonia, criterion validity was not assessed. Only questionnaires developed and published in English and studies or manuals written in English were eligible, thus, cross-cultural validity was not determined.

Methods

The PRISMA statement [12] and the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) [7, 8] 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 [8]; and 3) rating the quality of each measurement property for all questionnaires using criteria laid out by Terwee and colleagues [13] and Schellingerhout and colleagues [14].

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 or schooling, were excluded from this review. Single-item questionnaires or questionnaires that were not a comprehensive measure, were also excluded. For measures to be considered comprehensive, they should have had at least an overall or summative score. 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 Table 1). Next, additional searches were conducted to identify publications on the psychometric properties of the retrieved questionnaires (see Table 2). The final searches were conducted in November 2016. 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 [7]. 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 3 presents definitions for each measurement property used for this review, as guided by the COSMIN statement [7]. 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 [8] 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 to 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. Terwee and colleagues [15] introduced a 'worst rating counts' system in which the final quality rating for a measurement property was equivalent to the lowest rating given to any of the items contained in the checklist for that property. As this method impedes the detection of subtle differences in methodological quality between studies, a revised scoring procedure was developed [16-18], and this is the method used in this review. Final quality ratings for measurement properties are presented as a percentage using the following formula:

$$Total\ score\ per\ psychometric\ property = \frac{(Total\ score\ obtained - Min\ score\ possible)}{(Max\ score\ possible - Min\ score\ possible)} \times 100\%$$

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 the quality criteria per measurement property developed by Terwee and colleagues [13] and Schellingerhout and colleagues [14] (see Table 4). 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 the criteria set out by Schellingerhout and colleagues [14]. These criteria, or 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 (1,118 records) and PubMed (1,487 records) were identified. Figure 1 presents the flow diagram according to PRISMA [19]. A total of 67 questionnaires were assessed for eligibility, resulting in 15 questionnaires meeting all inclusion criteria. Table 5 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-seven 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 (EASE) [20], Glottal Function Index (GFI) [21], Singing Voice Handicap Index (SVHI) [22], Singing Voice Handicap Index-10 (SVHI-10) [23], Transgender Self-Evaluation Questionnaire (TSEQ) [24], Transsexual Voice Questionnaire - Male to Female (TVQ^{MtF}) [25], Vocal Fatigue Index (VFI) [26], Vocal Performance Questionnaire (VPQ) [27], Voice Capabilities Questionnaire (VCQ) [28], Voice Disability Coping Questionnaire (VDCQ) [29], Voice Handicap Index (VHI or VHI-30) [30], Voice Handicap Index-10 (VHI-10) [31], Voice Rating Scale (VRS) [32], Voice-Related Quality of Life (V-RQOL) [33] and Voice Symptom Scale (VoiSS) [34].

Details on the 47 studies on the development and validation of the included questionnaires on FHS and/or HR-QoL in dysphagia are summarised in Table 6. Table 7 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 [8] was used to assess the methodological quality of the 47 included studies. Table 8 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 one study received a poor COSMIN rating [35] for one of the analyses (hypothesis testing), and so it was excluded from further analysis. 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 (nine studies, mainly intra-rater reliability), content validity (12 studies) and structural validity (nine 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

Table 9 presents the quality of the psychometric properties retrieved from 47 included research articles for all 15 questionnaires based on the quality criteria by Terwee and colleagues [13] and Schellingerhout and colleagues [14]. Details on rating criteria are summarised in Table 4. The overall, integrated quality score for each measurement property per questionnaire was determined using the criteria Schellingerhout and colleagues [14], and is presented in Table 10. The overall level of psychometric quality is determined by integrating the methodological quality ratings of the included studies using the COSMIN checklist (Table 8) with the quality criteria for measurement properties of the questionnaires according to Terwee and colleagues [13] and Schellingerhout and colleagues [14] (Table 9).

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 47 studies describing at least one psychometric property of one or more of the included questionnaires. No manuals were retrieved. Twelve 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 was 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 indicates an incomplete psychometric overview and the generalisability and interpretation of results remain limited.

For two questionnaires, only data on a single psychometric characteristic was retrieved and for another three questionnaires data were found on two characteristics. For six questionnaires data was reported on for four psychometric characteristics, and for four questionnaires data was 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 [25, 28], 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^{MtF} (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^{MtF}). Even so, 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 [36]. Likewise, the use of a questionnaire with negative psychometric evidence cannot be justified based on its psychometric properties.

Francis et al. [11] summarised their findings on the psychometric properties of voice-related patient-reported measures based on their newly developed 18-item dichotomised checklist (present or absent) of evaluative criteria to operationalise measurement characteristics [9]. The COSMIN group criticised the methodological shortcomings of this checklist strongly [10], indicating that the psychometric properties of questionnaires in dysphonia remain unclear based on the review by Francis and colleagues. Given that the

inclusion and exclusion criteria of this review differed to that of Francis and colleagues, eight of the 15 included questionnaires in the current review were also reviewed by Francis and colleagues [11]: GFI, VFI, VPQ, VDCQ, VHI, VHI-10, V-RQOL and VoiSS. When considering the eight measures that overlap between both reviews, Francis and colleagues [11] 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 Francis and colleagues 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 by Francis and colleagues have methodological shortcomings as outlined by Terwee et al. [10], 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 according to Francis and colleagues [11] differ substantially from the psychometric data reported in the current review using the COSMIN taxonomy and checklist. Our findings contra-indicates the use of the simplified checklist by Francis and colleagues [9] 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). One single study by Phyland, Pallant, Benninger, Thibeault, Greenwood, Smtih and Vallance [20] 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 are 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 [37]. 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 [38]. 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, 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 47 studies reporting on psychometric characteristics of the included questionnaires. Quality criteria by Terwee and colleagues [13] and Schellingerhout and colleagues [14] were used to rate the psychometric data on measurement properties for each study. An overall quality score per measurement property per questionnaire was determined by applying the criteria or levels of evidence by Schellingerhout and colleagues [14]. 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 modelling and international consensus-based psychometric quality criteria and terminology, such as the COSMIN framework.

List of 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
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^{MtF}	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

Declarations

- Ethics approval and consent to participate: Not applicable
- Consent for publication: Not applicable
- Availability of data and material: All data generated or analysed during this study are included in this published article [and its supplementary information files.
- Competing interests: DP authored one of the instruments included in this review (EASE), however 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.
- Funding: No funding was received.
- Authors' contributions: RS and RC conceptualised the study. J-HK, KD, Y-WC, DD and DP extracted data from the literature. J-HK, KD, Y-WC and DD analysed the quality of studies and psychometric properties. All authors were major contributors in writing the manuscript and have read and approved the final manuscript.
- Acknowledgements: The authors express their gratitude to Grace Martin and Melinda Licciardello for providing support during the start-up of this review.

References

1. Cohen SM, Kim J, Roy N, Asche C, Courey M: **Prevalence and causes of dysphonia in a large treatment-seeking population.** *The Laryngoscope* 2012, **122:**343-348.
2. Roy N, Merrill RM, Gray SD, Smith EM: **Voice disorders in the general population: Prevalence, risk factors, and occupational impact.** *The Laryngoscope* 2005, **115:**1988-1995.
3. Roy N, Merrill RM, Thibeault S, Gray SD, Smith EM: **Voice disorders in teachers and the general population: Effects on work performance, attendance, and future career choices.** *JSLHR* 2004, **47:**542-551.
4. Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, Van de Heyning P, Woisard V: **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 Phoniatrics of the European Laryngological Society (ELS).** *Eur Arch Otorhinolaryngol* 2001, **258:**77-82.
5. Speyer R: **Oropharyngeal dysphagia: Screening and assessment.** In *Dysphagia Otolaryngologic Clinics of North America. Volume* 46. Edited by Altman KW. The Clinics: Internal Medicine: Elsevier Health Sciences; 2013: 989-1008
6. Ferrans CE, Zerwic JJ, Wilbur JE, Larson JL: **Conceptual model of health-related quality of life.** *Journal of Nursing Scholarship* 2005, **37:**336-342.
7. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC: **The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes.** *Journal of Clinical Epidemiology* 2010, **63:**737-745.
8. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC: **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* 2010, **19:**539-549.
9. Francis DO, McPheeters ML, Noud M, Penson DF, Feurer ID: **Checklist to operationalize measurement characteristics of patient-reported outcome measures.** *Systematic Reviews* 2016, **5:**129.
10. **Comment on "Checklist to operationalize measurement characteristics of patient-reported outcome measures"** [http://www.cosmin.nl/commentory-on-francis-do-syst-revs-.html]
11. Francis DO, Daniero JJ, Hovis KL, Sathe N, Jacobson B, Penson DF, Feurer ID, McPheeters ML: **Voice-related patient-reported outcome measures: A systematic review of instrument development and validation.** *JSLHR* 2017, **60:**62-88.
12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D: **The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration.** *Annals of Internal Medicine* 2009, **151:**W1-W30.
13. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC: **Quality criteria were proposed for measurement properties of health status questionnaires.** *Journal of Clinical Epidemiology* 2007, **60:**34-42.
14. Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, Henrica C, Terwee CB: **Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review.** *Quality of Life Research* 2012, **21:**659-670.
15. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC: **Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist.** *Quality of Life Research* 2012, **21:**651-657.
16. Speyer R, Cordier R, Kertscher B, Heijnen BJ: **Psychometric properties of questionnaires on functional health status in oropharyngeal dysphagia: a systematic literature review.** *BioMed Research International* 2014.

17. Cordier R, Speyer R, Chen YW, Wilkes-Gillan S, Brown T, Bourke-Taylor H, Doma K, Leicht A: **Evaluating the Psychometric Quality of Social Skills Measures: A Systematic Review.** *PLoS ONE* 2015, **10:**e0132299.
18. Cordier R, Chen Y-W, Speyer R, Totino R, Doma K, Leicht A, Brown N, Cuomo B: **Child-report measures of Occupational Performance: A systematic review.** *PLoS ONE* 2016, **11:**e0147751.
19. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P: **Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.** *Annals of Internal Medicine* 2009, **151:**264-269.
20. Phyland DJ, Pallant JF, Benninger MS, Thibeault SL, Greenwood KM, Smith JA, Vallance N: **Development and preliminary validation of the EASE: A tool to measure perceived singing voice function.** *Journal of Voice* 2013, **27:**454-462.
21. Bach KK, Belafsky PC, Wasylik K, Postma GN, Koufman JA: **Validity and reliability of the glottal function index.** *Archives of Otolaryngology–Head & Neck Surgery* 2005, **131:**961-964.
22. Cohen SM, Jacobson BH, Garrett CG, Noordzij JP, Stewart MG, Attia A, Ossoff RH, Cleveland TF: **Creation and validation of the Singing Voice Handicap Index.** *Annals of Otology, Rhinology & Laryngology* 2007, **116:**402-406.
23. Cohen SM, Statham M, Rosen CA, Zullo T: **Development and validation of the singing voice handicap-10.** *The Laryngoscope* 2009, **119:**1864-1869.
24. Davis S: **Transgender Self-Evaluation Questionnaire.** In *Voice and communication therapy for the transgender/transsexual client: A comprehensive clinical guide.* Edited by Adler H, Mordaunt M. San Diego: Plural Publishing; 2006: 485-487
25. Dacakis G, Davies S, Oates JM, Douglas JM, Johnston JR: **Development and preliminary evaluation of the transsexual voice questionnaire for male-to-female transsexuals.** *Journal of Voice* 2013, **27:**312-320.
26. Nanjundeswaran C, Jacobson BH, Gartner-Schmidt J, Verdolini Abbott K: **Vocal Fatigue Index (VFI): Development and Validation.** *J Voice* 2015, **29:**433-440.
27. Carding PN, Horsley IA: **An evaluation study of voice therapy in non‐organic dysphonia.** *International Journal of Language & Communication Disorders* 1992, **27:**137-158.
28. Buckley KL, O'Halloran PD, Oates JM: **Occupational Vocal Health of Elite Sports Coaches: An Exploratory Pilot Study of Football Coaches.** *Journal of Voice* 2015, **29:**476-483.
29. Epstein R, Hirani SP, Stygall J, Newman SP: **How Do Individuals Cope With Voice Disorders? Introducing the Voice Disability Coping Questionnaire.** *Journal of Voice* 2009, **23:**209-217.
30. Jacobson BH, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger MS, Newman CW: **The voice handicap index (VHI) development and validation.** *American Journal of Speech-Language Pathology* 1997, **6:**66-70.
31. Rosen CA, Lee AS, Osborne J, Zullo T, Murry T: **Development and Validation of the Voice Handicap Index-10.** *The Laryngoscope* 2004, **114:**1549-1556.
32. Wingate JM, Brown WS, Shrivastav R, Davenport P, Sapienza CM: **Treatment Outcomes for Professional Voice Users.** *Journal of Voice* 2007, **21:**433-449.
33. Hogikyan ND, Sethuraman G: **Validation of an instrument to measure voice-related quality of life (V-RQOL).** *Journal of Voice* 1999, **13:**557-569.
34. Deary IJ, Wilson JA, Carding PN, MacKenzie K: **VoiSS: A patient-derived Voice Symptom Scale.** *Journal of Psychosomatic Research* 2003, **54:**483-489.
35. Wilson JA, Webb A, Carding PN, Steen IN, MacKenzie K, Deary IJ: **The Voice Symptom Scale (VoiSS) and the Vocal Handicap Index (VHI): a comparison of structure and content.** *Clin Otolaryngol Allied Sci* 2004, **29:**169-174.
36. Mokkink LB, Terwee CB, Knol DL, Stratford PW, J. A, D.L. P, L.M. B, de Vet HCW: **The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content.** *BMC Med Res Methodol* 2010, **10**.

37. Linacre JM: *Rasch-model computer programs: Program manual 3.92.0.* Chicago, IL: Mesa-Press; 2016.
38. Reise S, Henson J: **A discussion of modern versus traditional psychometrics as applied to personality assessment scales.** 2003, **81:**93-103.
39. Brutten G, Vanryckeghem M: *Behavior Assessment Battery: A multi-dimensional and evidence-based approach to diagnostic and therapeutic decision making for children who stutter.* The Netherlands: Belgium & Acco Publishers; 2003.
40. Borg G: **Psychological scaling with applications in physical work and the perception of exertion.** *Scan J Work Environ Health* 1990, **16(Suppl.1):**55-58.
41. Ford Baldner E, Doll E, van Mersbergen MR: **A review of measures of vocal effort with a preliminary study on the establishment of a vocal effort measure.** *Journal of Voice* 2015, **29:**530-541.
42. Ricci-Maccarini A, De Maio, V., Murry, T., & Schindler, A.: **Development and validation of the children's voice handicap index-10 (CVHI-10).** *Journal of Voice* 2013, **27:**258-e223.
43. Fussi F: **La valutazione del canto.** In *La voce del cantante.* Edited by Fussi F. Torino: Omega; 2005: 33-68
44. Baylor C, Yorkston KM, Eadie TL, Jiseon K, Hyewon C, Amtmann D: **The Communicative Participation Item Bank (CPIB): Item Bank Calibration and Development of a Disorder-Generic Short Form.** *Journal of Speech, Language & Hearing Research* 2013, **56:**1190-1208.
45. Giannini SP, Latorre Mdo R, Ferreira LP: **Condition of Vocal Production-Teacher questionnaire: comparison of responses on Likert scale and visual analog scale.** *Codas* 2016, **28:**53-58.
46. Hu A, Isetti D, Hillel AD, Waugh P, Comstock B, Meyer TK: **Disease-Specific Self-Efficacy in Spasmodic Dysphonia Patients.** *Otolaryngology–Head and Neck Surgery* 2013, **148:**450-455.
47. Nemr K, Simões-Zenari M, Duarte JMDT, Lobrigate KE, Bagatini FA: **Dysphonia risk screening protocol.** *Clinics* 2016, **71:**114-127.
48. Wuyts FL, De Bodt M, Molenbergths G, Remacle M, Heylen L, Millet B, Van Lierde K, Raes J, Van de Heyning P: **The Dysphonia Severity Indesc: An objective measure of vocal quality based on a multiparameter approach.** *Journal of Speech Language and Hearing Research* 2000, **43:**796-809.
49. Robinson K, Gatehouse S, Browning GG: **Measuring Patient Benefit from Otorhinolaryngological Surgery and Therapy.** *The Annals of Otology, Rhinology & Laryngology* 1996, **105:**415-422.
50. Kubba H, Swan IRC, Gatehouse S: **The Glasgow Children's Benefit Inventory: A New Instrument for Assessing Health-Related Benefit after an Intervention.** *Annals of Otology, Rhinology & Laryngology* 2004, **113:**980-986.
51. Goldman SL, Hargrave J, Hillman RE, Holmberg E, Gress C: **Stress, Anxiety, Somatic Complaints, and Voice Use in Women With Vocal Nodules: Preliminary Findings.** *American Journal of Speech-Language Pathology* 1996, **5:**44-54.
52. Karnell MP, Melton SD, Childes JM, Coleman TC, Dailey SA, Hoffman HT: **Reliability of Clinician-Based (GRBAS and CAPE-V) and Patient-Based (V-RQOL and IPVI) Documentation of Voice Disorders.** *Journal of Voice* 2007, **21:**576-590.
53. Dehqan A, Yadegari F, Asgari A, Scherer RC, Dabirmoghadam P: **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* 2017, **31** 113-119.
54. Baylor C, Yorkston K, Eadie T, Miller R, Amtmann D: **Levels of speech Usage: A self-report scale for describing how people use speech.** *J Med Speech Lang Pathol* 2008, **16:**191-198.
55. Llewellyn-Thomas HA, Sutherland HJ, Hogg SA, Ciampi A, Harwood AR, Keane TJ, Till JE, Boyd NF: **Linear analogue self-assessment of voice quality in laryngeal cancer.** *Journal of chronic diseases* 1984, **37:**917-924.

56. Hartelius L, Elmberg M, Holm R, Lövberg AS, Nikolaidis S: **Living with Dysarthria: Evaluation of a Self-Report Questionnaire.** *Folia Phoniatrica et Logopaedica* 2008, **60:**11-19.
57. Costa FP, Diaferia G, Behlau M: **Communicative aspects and coping strategies in patients with Parkinson's disease.** *Codas* 2016, **28:**46-52.
58. Priebe S, Huxley P, Knight S, Evans S: **Application and Results of the Manchester Short Assessment of Quality of Life (Mansa).** *International Journal of Social Psychiatry* 1999, **45:**7-12.
59. Moreti F, Rocha C, Borrego MCM, Behlau M: **Desvantagem vocal no canto: análise do protocolo Índice de Desvantagem para o Canto Moderno – IDCM.** *Rev Soc Bras Fonoaudiol* 2011, **16:**146-151.
60. Koufman JA, Isaacson G: **The spectrum of vocal dysfunction.** *Otolaryngologic Clinics of North America* 1991, **24:**985-988.
61. Zur KB, Cotton S, Kelchner L, Baker S, Weinrich B, Lee L: **Pediatric Voice Handicap Index (pVHI): A new tool for evaluating pediatric dysphonia.** *International Journal of Pediatric Otorhinolaryngology* 2007, **71:**77-82.
62. Hartnick CJ: **Validation of a pediatric voice quality-of-life instrument: The pediatric voice outcome survey.** *Archives of Otolaryngology–Head & Neck Surgery* 2002, **128:**919-922.
63. Boseley ME, Cunningham MJ, Volk MS, Hartnick CJ: **Validation of the pediatric voice-related quality-of-life survey.** *Archives of Otolaryngology–Head & Neck Surgery* 2006, **132:**717-720.
64. Smith E, Verdolini K, Gray S, Nichols S, Lemke J, Barkmeier J, Dove H, Hoffman H: **Effect of voice disorders on quality of life.** *Journal of Medical Speech-Language Pathology* 1996, **4:**223-244.
65. De Araujo Pernambuco L, Espelt A, Virgilio Magalhaes Junior H, Veiga Andersen Cavalcanti R, Costa de Lima K: **Screening for voice disorders in older adults - Part I: Validity evidence based on test content and response processes.** *J Voice* 2016, **30:**246.e249-246.e217.
66. Vitali C, Borghi E, Napoletano A, Polini F, Caronni M, Ammenti P, Cattaneo D: **Oropharyngolaryngeal disorders in scleroderma: Development and validation of the SLS scale.** *Dysphagia* 2010, **25:**127-138.
67. Harris T, Collins S, Clarke DD: **An interdisciplinary voice clinic.** In *Voice disorders and their management.* Edited by Fawcus M. London: Croom Helm; 1986: 283-292
68. Sheehan DV: *The Anxiety Disease.* New York (NY): Scribner; 1983.
69. Garcia GM, Lopez JM, Lara EYM: **Voice habits and behaviours: Voice care among Flamenco singers.** *J Voice* 2017, **31:**246.e211-e219.
70. Jones SM, Carding PN, Drinnan MJ: **Exploring the relationship between severity of dysphonia and voice-related quality of life.** *Clinical Otolaryngology* 2006, **31:**411-417.
71. Epstein R, Stygall J, Newman S: **The short-term impact of Botox injections on speech disability in adductor spasmodic dysphonia.** *Disability and Rehabilitation* 1997, **19:**20-25.
72. Isetti D, Meyer T: **Workplace Productivity and Voice Disorders: A Cognitive Interviewing Study on Presenteeism in Individuals With Spasmodic Dysphonia.** *Journal of Voice* 2014, **28:**700-710.
73. Hill DS, Akhtar S, Corroll A, Croft CB: **Quality of life issues in recurrent respiratory papillomatosis.** *Clinical Otolaryngology & Allied Sciences* 2000, **25:**153-160.
74. Kirsh ER, van Leer E, Phero HJ, Xie C, Khosla S: **Factors assoicated with singers' perceptions of choral singing well-being.** *Journal of Voice* 2013, **27:**786.e725-786.e732.
75. Nam I, Bae J, Shim M, Hwang Y, Kim M, Sun D: **The importance of preoperative laryngeal examination before thyroidectomy and the usefulness of a voice questionnaire in screening.** *World Journal of Surgery* 2012, **36:**303-309.
76. Stewart CF, Allen EL, Tureen P, Diamond BE, Blitzer A, Brin MF: **Adductor spasmodic dysphonia: standard evaluation of symptoms and severity.** *Journal of Voice* 1997, **11:**95-103.

77. Teixeira LC, Rodrigues ALV, Silva ÁFGD, Azevedo R, Gama ACC, Behlau M: **The use of the URICA-VOICE questionnaire to identify the stages of adherence to voice treatment.** *Codas* 2013, **25:**8-15.
78. Speyer R, Wieneke GH, Dejonckere PH: **Self‐assessment of voice therapy for chronic dysphonia.** *Clinical Otolaryngology* 2004, **29:**66-74.
79. Paolillo NP, Pantaleo G: **Development and Validation of the Voice Fatigue Handicap Questionnaire (VFHQ): Clinical, Psychometric, and Psychosocial Facets.** *Journal of Voice* 2015, **29:**91-100.
80. Mathieson L, Hirani SP, Epstein R, Baken RJ, Wood G, Rubin JS: **Laryngeal Manual Therapy: A Preliminary Study to Examine its Treatment Effects in the Management of Muscle Tension Dysphonia.** *Journal of Voice* 2009, **23:**353-366.
81. Ma EP, Yiu EM: **Voice Activity and Participation Profile: assessing the impact of voice disorders on daily activities.** *Journal of Speech, Language & Hearing Research* 2001, **44:**511-524.
82. Konnai RM, Jayaram M, Scherer RC: **Development and Validation of a Voice Disorder Outcome Profile for an Indian Population.** *Journal of Voice* 2010, **24:**206-220.
83. Gugatschka M, Rechenmacher J, Chibidziura J, Friedrich G: **Vergleichbarkeit und Urechnug von Stimmstorungsindex (SSI) und Voice Handicap Index (VHI).** *Laryngo-Rhino-Otologie* 2007, **86:**785.
84. Li H, Huang Z, Hu R, Zhang L, Xu W: **Study on the simplified Chinese version of the voice handicap index.** *Journal of Voice* 2012, **26:**365-371.
85. Zraick RI, Risner BY, Smith-Olinde L, Gregg BA, Johnson FL, McWeeny EK: **Patient Versus Partner Perception of Voice Handicap.** *Journal of Voice* 2007, **21:**485-494.
86. Lyberg-Åhlander V, Rydell R, Eriksson J, Schalén L: **Throat related symptoms and voice: development of an instrument for self assessment of throat-problems.** *Ear, Nose and Throat Disorders* 2010, **10:**5.
87. El-Banna M, Ras YA: **Formulation of a voice problem self assessment scale (VPSS) as a patient based tool for Egyptian dysphonic population.** *Alexandria Journal of Medicine* 2010, **46:**119-126.
88. Gillespie AI, Abbott KV: **The influence of clinical terminology on self-efficacy for voice.** *Logopedics Phoniatrics Vocology* 2011, **36:**91-99.
89. Gliklich RE, Glovsky RM, Montgomery WW: **Validation of a Voice Outcome Survey for Unilateral Vocal Cord Paralysis.** *Otolaryngology–Head and Neck Surgery* 1999, **120:**153-158.
90. Reilly MC, Zbrozek AS, Dukes EM: **The validity and reproducibility of a work productivity and activity impairment instrument.** *Pharmacoeconomics* 1993, **4:**353-365.
91. Rosow DE, Szczupak M, Saint-Victor S, Gerhard JD, Dupont C, Lo K: **The economic impact of vocal attrition in public school teachers in Miami-Dade County.** *Laryngoscope* 2016, **126:**665-671.
92. Phyland DJ, Pallant JF, Thibeault SL, Benninger MS, Vallance N, Smith JA: **Measuring vocal function in professional music theater singers: construct validation of the Evaluation of the Ability to Sing Easily (EASE).** *Folia Phoniatrica et Logopaedica* 2014, **66:**100-108.
93. Buckmire RA, Bryson PC, Patel MR: **Type I gore-tex laryngoplasty for glottic incompetence in mobile vocal folds.** *J Voice* 2011, **25:**288-292.
94. Castelblanco L, Habib M, Stein DJ, de Quadros A, Cohen SM, Noordzij JP: **Singing Voice Handicap and videostrobolaryngoscopy in healthy professional singers.** *Journal of Voice* 2014, **28:**608-613.
95. Hancock AB, Krissinger J, Owen K: **Voice perceptions and quality of life of transgender people.** *J Voice* 2011, **25:**553-558.
96. Hancock AB: **An ICF Perspective on Voice-related Quality of Life of American Transgender Women.** *J Voice* 2017, **31:**115 e111-115 e118.

97. Deary IJ, Webb A, Mackenzie K, Wilson JA, Carding PN: **Short, self-report voice symptom scales: psychometric characteristics of the voice handicap index-10 and the vocal performance questionnaire.** *Otolaryngol Head Neck Surg* 2004, **131:**232-235.
98. Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen IN, Wilson JA: **Optimising outcome assessment of voice interventions, I: Reliability and validity of three self-reported scales.** *Journal of Laryngology and Otology* 2007, **121:**763-767.
99. Awan SN, Roy N, Cohen SM: **Exploring the relationship between spectral and cepstral measures of voice and the Voice Handicap Index (VHI).** *J Voice* 2014, **28:**430-439.
100. Elam JC, Ishman SL, Dunbar KB, Clarke JO, Gourin CG: **The relationship between depressive symptoms and Voice Handicap Index scores in laryngopharyngeal reflux.** *Laryngoscope* 2010, **120:**1900-1903.
101. Fulljames N, Harris S: **Voice outcome measures: correlations with patients' assessment of their condition and the effectiveness of voice therapy.** *Logoped Phoniatr Vocol* 2006, **31:**23-35.
102. Kazi R, De Cordova J, Singh A, Venkitaraman R, Nutting CM, Clarke P, Rhys-Evans P, Harrington KJ: **Voice-related Quality of Life in Laryngectomees: Assessment Using the VHI and V-RQOL Symptom Scales.** *Journal of Voice* 2007, **21:**728-734.
103. Portone CR, Hapner ER, McGregor L, Otto K, Johns MM, 3rd: **Correlation of the Voice Handicap Index (VHI) and the Voice-Related Quality of Life Measure (V-RQOL).** *J Voice* 2007, **21:**723-727.
104. Stomeo F, Tosin E, Morolli F, Bianchini C, Ciorba A, Pastore A, Pelucchi S: **Comparison of subjective and objective tools in transoral laser cordectomy for early glottic cancer: importance of voice handicap index.** *Int J Immunopathol Pharmacol* 2013, **26:**445-451.
105. Wheeler KM, Collins SP, Sapienza CM: **The relationship between VHI scores and specific acoustic measures of mildly disordered voice production.** *J Voice* 2006, **20:**308-317.
106. Childs LF, Bielinski C, Toles L, Hamilton A, Deane J, Mau T: **Relationship between patient-perceived vocal handicap and clinician-rated level of vocal dysfunction.** *Laryngoscope* 2015, **125:**180-185.
107. Davis KM, Sandage MJ, Plexico L, Pascoe DD: **The Perception of Benefit of Vocalization on Sport Performance When Producing Maximum Effort.** *J Voice* 2016, **30:**639 e611-636.
108. Eadie TL, Lamvik K, Baylor CR, Yorkston KM, Kim J, Amtmann D: **Communicative participation and quality of life in head and neck cancer.** *Ann Otol Rhinol Laryngol* 2014, **123:**257-264.
109. Gillespie AI, Gooding W, Rosen C, Gartner-Schmidt J: **Correlation of VHI-10 to voice laboratory measurements across five common voice disorders.** *J Voice* 2014, **28:**440-448.
110. Hu A, Hillel A, Meyer T: **Factors Associated With Patient-perceived Hoarseness in Spasmodic Dysphonia Patients.** *J Voice* 2016, **30:**769 e723-769 e726.
111. Kupfer RA, Cadalli Tatar E, Barry JO, Allen CT, Merati AL: **Anatomic Derkay Score Is Associated with Voice Handicap in Laryngeal Papillomatosis in Adults.** *Otolaryngol Head Neck Surg* 2016, **154:**689-692.
112. Nichols B, Bock JM, Blumin JH: **Dysphonia in nursing home and assisted living residents: prevalence and association with frailty.** *J Voice* 2015, **29:**79-82.
113. Romak JJ, Orbelo DM, Maragos NE, Ekbom DC: **Correlation of the Voice Handicap Index-10 (VHI-10) and Voice-Related Quality of Life (V-RQOL) in patients with dysphonia.** *J Voice* 2014, **28:**237-240.
114. Stachler RJ, Schultz LR, Nerenz D, Yaremchuk KL: **PROMIS evaluation for head and neck cancer patients: a comprehensive quality-of-life outcomes assessment tool.** *Laryngoscope* 2014, **124:**1368-1376.
115. Willis J, Michael DD, Boyer H, Misono S: **Prevalence and Severity of Dysphonia in Patients with Cystic Fibrosis: A Pilot Study.** *Otolaryngol Head Neck Surg* 2015, **153:**88-93.

116. Bornbaum CC, Day AM, Doyle PC: **Examining the construct validity of the V-RQOL in speakers who use alaryngeal voice.** *Am J Speech Lang Pathol* 2014, **23:**196-202.
117. Kupfer RA, Hogikyan EM, Hogikyan ND: **Establishment of a normative database for the Voice-Related Quality of Life (V-RQOL) measure.** *J Voice* 2014, **28:**449-451.
118. Murry T, Medrado R, Hogikyan ND, Aviv JE: **The relationship between ratings of voice quality and quality of life measures.** *J Voice* 2004, **18:**183-192.
119. Tanner K, Pierce JL, Merrill RM, Miller KL, Kendall KA, Roy N: **The Quality of Life Burden Associated With Voice Disorders in Sjogren's Syndrome.** *Ann Otol Rhinol Laryngol* 2015, **124:**721-727.
120. Montgomery J, Hendry J, Wilson JA, Deary IJ, MacKenzie K: **Pragmatic detection of anxiety and depression in a prospective cohort of voice outpatient clinic attenders.** *Clin Otolaryngol* 2016, **41:**2-7.

Table 1. *Search strategies per literature database to retrieve questionnaires on FHS and/or HR-QoL.*

	Databases and Search terms	**Limits**	**Number of records**
Subject Headings	**Embase:** (questionnaire/ OR “severity of illness index”/ OR health status/ OR health survey/) AND (dysphonia/ OR voice disorder/ OR aphonia/ OR singing/)	N/A	846
	PubMed: ("Surveys and Questionnaires"[Mesh] OR "Severity of illness index"[mesh] OR "health Status"[Mesh]) AND (“Dysphonia” [Mesh] OR “Voice Disorders” [Mesh] OR “Hoarseness” [Mesh] OR “Aphonia” [Mesh] OR “Singing” [Mesh])	N/A	1,284
Free Text	**Embase:** (questionnaire* OR survey*) AND (dysphon* OR hoarseness OR roughness OR voice OR aphon* OR sing* or sang)	Time limit: 2016 to Current	331
	PubMed: *As per Embase Free Text*	Time limit: 01/01/2016 to 22/11/2016	281

Table 2. *Database search strategies to retrieve publications on the psychometric properties of the retrieved questionnaires.*

	Databases and Search terms	**Limits**	**Number of records**
Subject Headings	**Embase**: ((Evaluation of the Ability to Sing Easily) OR EASE OR (Glottal Function Index) OR (GFI) OR (Singing Voice Handicap Index) OR SVHI OR (Singing Voice Handicap Index-10) OR SVHI-10 OR (Transgender Self-Evaluation Questionnaire) OR TSEQ OR (Transsexual Voice Questionnaire - Male to Female) OR TVQ-MtF OR (Voice Index) OR (Vocal Fatigue Index) OR VFI OR (Vocal Performance Questionnaire) OR VPQ OR (Voice Capabilities Questionnaire) OR VCQ OR (Voice Disability Coping Questionnaire) OR VDCQ OR (Voice Handicap Index) OR VHI OR (Voice Handicap Index-10) OR VHI-10 OR (Voice Outcome Survey) OR (Voice Rating Scale) OR VRS OR (Voice-Related Quality of Life) OR V-RQOL OR (Voice Symptom Scale) OR VoiSS) AND (psychometry/ OR validity/ OR reliability/ OR measurement error/ OR measurement precision/ OR measurement repeatability/ OR error/ OR statistical bias/ OR test retest reliability/ OR intrarater reliability/ OR interrater reliability/ OR accuracy/ OR criterion validity/ OR internal validity/ OR face validity/ OR external validity/ OR discriminant validity/ OR concurrent validity/ OR qualitative validity/ OR construct validity/ OR content validity/)	NA	339
	PubMed: ((Evaluation of the Ability to Sing Easily) OR EASE OR (Glottal Function Index) OR (GFI) OR (Singing Voice Handicap Index) OR SVHI OR (Singing Voice Handicap Index-10) OR SVHI-10 OR (Transgender Self-Evaluation Questionnaire) OR TSEQ OR (Transsexual Voice Questionnaire - Male to Female) OR TVQ-MtF OR (Voice Index) OR (Vocal Fatigue Index) OR VFI OR (Vocal Performance Questionnaire) OR VPQ OR (Voice Capabilities Questionnaire) OR VCQ OR (Voice Disability Coping Questionnaire) OR VDCQ OR (Voice Handicap Index) OR VHI OR (Voice Handicap Index-10) OR VHI-10 OR (Voice Outcome Survey) OR (Voice Rating Scale) OR VRS OR (Voice-Related Quality of Life) OR V-RQOL OR (Voice Symptom Scale) OR VoiSS) AND ("Psychometrics"[Mesh] OR "Reproducibility of Results"[Mesh] OR "Validation Studies as Topic"[Mesh] OR "Validation Studies" [Publication Type] OR "Bias (Epidemiology)"[Mesh] OR "Observer Variation"[Mesh] OR "Selection Bias"[Mesh] OR "Diagnostic Errors"[Mesh] OR "Dimensional Measurement Accuracy"[Mesh] OR “Predictive Value of Tests"[Mesh] OR "Discriminant Analysis"[Mesh])	NA	676

	Databases and Search terms	Limits	Number of records
Free Text	**Embase**: ((Evaluation of the Ability to Sing Easily) OR EASE OR (Glottal Function Index) OR (GFI) OR (Singing Voice Handicap Index) OR SVHI OR (Singing Voice Handicap Index-10) OR SVHI-10 OR (Transgender Self-Evaluation Questionnaire) OR TSEQ OR (Transsexual Voice Questionnaire - Male to Female) OR TVQ-MtF OR (Voice Index) OR (Vocal Fatigue Index) OR VFI OR (Vocal Performance Questionnaire) OR VPQ OR (Voice Capabilities Questionnaire) OR VCQ OR (Voice Disability Coping Questionnaire) OR VDCQ OR (Voice Handicap Index) OR VHI OR (Voice Handicap Index-10) OR VHI-10 OR (Voice Outcome Survey) OR (Voice Rating Scale) OR VRS OR (Voice-Related Quality of Life) OR V-RQOL OR (Voice Symptom Scale) OR VoiSS) AND (psychometric* OR reliabilit* OR validit* OR reproducibilit* OR bias).ti,ab.)	Publication date: 2016 – Current	33
	PubMed: ((Evaluation of the Ability to Sing Easily) OR EASE OR (Glottal Function Index) OR (GFI) OR (Singing Voice Handicap Index) OR SVHI OR (Singing Voice Handicap Index-10) OR SVHI-10 OR (Transgender Self-Evaluation Questionnaire) OR TSEQ OR (Transsexual Voice Questionnaire - Male to Female) OR TVQ-MtF OR (Voice Index) OR (Vocal Fatigue Index) OR VFI OR (Vocal Performance Questionnaire) OR VPQ OR (Voice Capabilities Questionnaire) OR VCQ OR (Voice Disability Coping Questionnaire) OR VDCQ OR (Voice Handicap Index) OR VHI OR (Voice Handicap Index-10) OR VHI-10 OR (Voice Outcome Survey) OR (Voice Rating Scale) OR VRS OR (Voice-Related Quality of Life) OR V-RQOL OR (Voice Symptom Scale) OR VoiSS) AND (psychometric*[Title/Abstract] OR reliabilit*[Title/Abstract] OR validit*[Title/Abstract] OR reproducibilit*[Title/Abstract] OR bias[Title/Abstract])	Publication date: 2016/01/01 to 2016/11/22	73

Table 3. *Definitions of measurement properties for Health-Related Patient-Reported Outcomes instruments according to COSMIN [7].*

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 are an adequate reflection of the performance of the items of the original version of the instrument*
	Criterion validity *Degree to which the scores of an instrument are an adequate reflection of a 'gold standard'*	
Responsiveness *Ability of an instrument to detect change over time in the construct to be measured*		
Interpretability[a] *Degree to which one can assign qualitative meaning to an instrument's quantitative scores or change in scores*		

[a] Interpretability is not considered a psychometric property.

Table 4. *Quality criteria for psychometric properties based on Terwee and colleagues [13] and Schellingerhout and colleagues [14]*

Property	**Score**[a]	**Quality criteria**[b]
Internal consistency	+	Factor analyses performed on adequate sample size (7 * # items and ≥ 100) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95; IRT[e]
	?	No factor analysis OR doubtful design or method
	-	Cronbach's alpha(s) <0.70 or >0.95, despite adequate design and method; IRT[e]
	±	Conflicting results
	NR	No information found on internal consistency
	NE	Not evaluated
Reliability (inter-rater reliability, intra-rater reliability, repeated measurement)	+	ICC or weighted Kappa ≥ 0.70
	?	Doubtful design or method (e.g., time interval not mentioned)
	-	ICC or weighted Kappa < 0.70, despite adequate design and method
	±	Conflicting results
	NR	No information found on reliability
	NE	Not evaluated
Measurement error[c]	+	MIC < SDC OR MIC outside the LOA OR convincing arguments that agreement is acceptable
	?	Doubtful design or method OR (MIC not defined AND no convincing arguments that agreement is acceptable)
	-	MIC ≥ SDC OR MIC equals or inside LOA, despite adequate design and method
	±	Conflicting results
	NR	No information found on agreement
	NE	Not evaluated
Content validity	+	A clear description is provided of the measurement aim, the target population, the concepts that are being measured, and the item selection AND target population and (investigators OR experts) were involved in item selection
	?	A clear description of above-mentioned aspects is lacking OR only target population involved OR doubtful design or method
	-	No target population involvement
	±	Conflicting results
	NR	No information found on target population involvement
	NE	Not evaluated
Structural validity	+	Factors should explain at least 50% of the variance; IRT[f]
	?	Explained variance not mentioned
	-	Factors explain < 50% of the variance; IRT[f]

Property	Score[a]	Quality criteria[b]
	±	Conflicting results
	NR	No information found on structural validity
	NE	Not evaluated
Hypothesis testing[d]	+	Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses
	?	Doubtful design or method (e.g., no hypotheses)
	-	Less than 75% of hypotheses were confirmed, despite adequate design and methods
	±	Conflicting results between studies within the same manual
	NR	No information found on hypotheses testing
	NE	Not evaluated
Criterion validity	+	Convincing arguments that gold standard is “gold” AND correlation with gold standard ≥0.70
	?	No convincing arguments that gold standard is “gold” OR doubtful design or method
	-	Correlation with gold standard <0.70, despite adequate design and method
	±	Conflicting results
	NR	No information found on criterion validity
	NE	Not evaluated
Responsiveness	+	SDC or SDC < MIC OR MIC outside the LOA OR RR > 1.96 OR AUC ≥ 0.70
	?	Doubtful design or method
	-	SDC or SDC ≥ MIC OR MIC equals or inside LOA OR RR ≤ 1.96 OR AUC < 0.70, despite adequate design and methods
	±	Conflicting results
	NR	No information found on responsiveness
	NE	Not evaluated

[a] Scores: + = positive rating; ? = indeterminate rating; - = negative rating; ± = conflicting data; NR = not reported; NE = not evaluated (for study of poor methodological quality according to COSMIN rating; data are excluded from further evaluation).
[b] Doubtful design or method = no clear description of the design or methods of the study, sample size smaller than 50 subjects (should be at least 50 in every (subgroup) analysis), or any important methodological weakness in the design or execution of the study.
[c] Measurement error: MIC = minimal important change; SDC = smallest detectable change; LOA = limits of agreement.
[d] Hypothesis testing: all correlations should be statistically significant (if not, these hypotheses are not confirmed) AND these correlations should be at least moderate (r > 0.5).
[e] For IRT: + Person Separation Index > 2 OR person reliability values ≥0.8; - Person Separation Index ≤2 OR person reliability values <0.8
[f] For IRT: + 1) Principal Component Analysis (PCA) of residuals: >60% of the variance explained by Rasch factor OR 2) Eigenvalue <3 on 1st contrast OR 1st contrast explains < 10% of the variance OR 3) (Overall item and person infit and outfit fit reported as a MnSq >0.7 or < 1.4, and outfit Z-STD values < ±2) AND (positive point-biserial correlations); - If none of the criteria 1) to 3) are met.

Figure 1. *Flow diagram of the reviewing process according to PRISMA. Adapted from Moher and colleagues [19]*

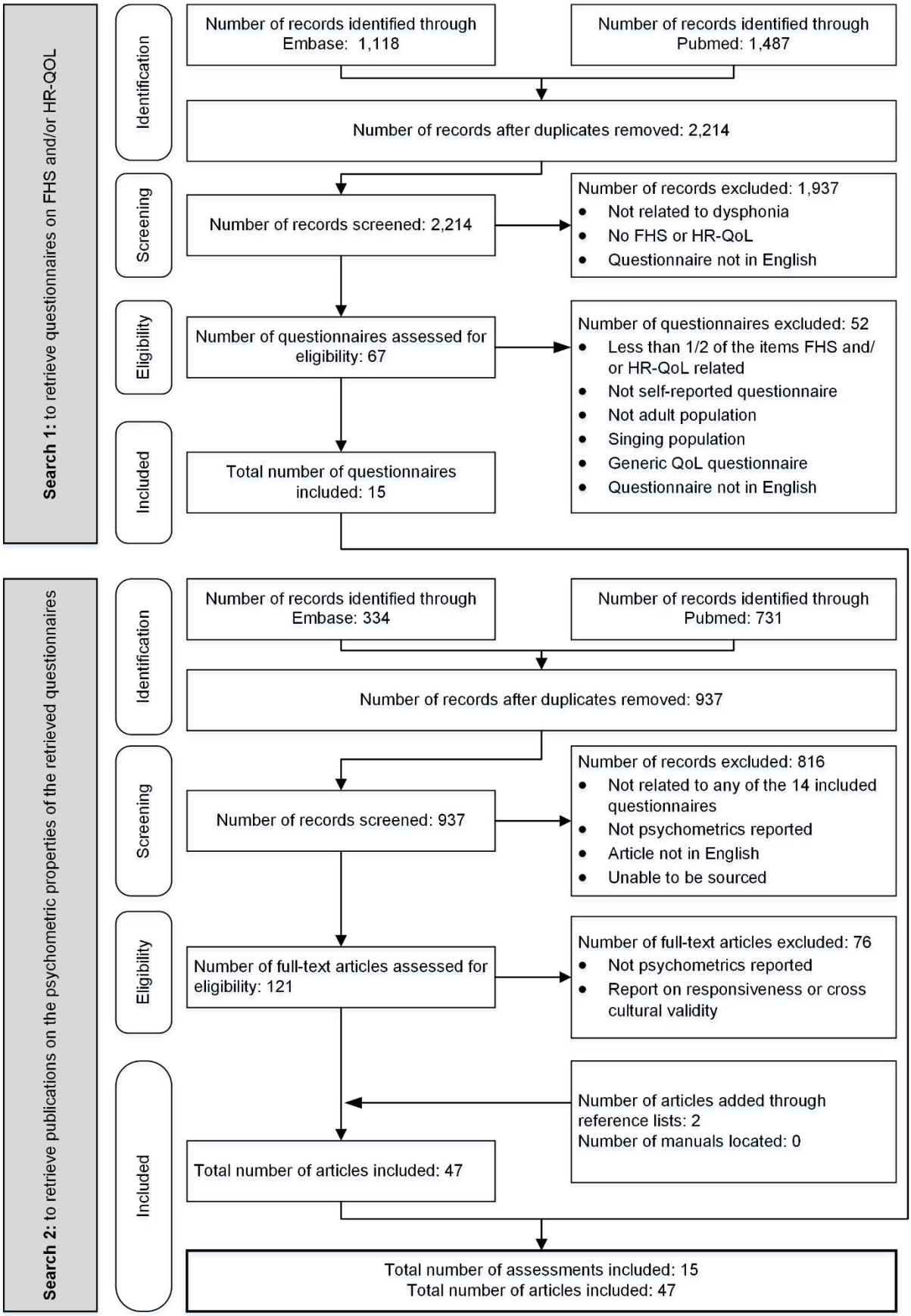
Search 1: to retrieve questionnaires on FHS and/or HR-QOL
Identification
Number of records identified through Embase: 1,118
Number of records identified through Pubmed: 1,487
Number of records after duplicates removed: 2,214
Screening
Number of records screened: 2,214
Number of records excluded: 1,937
Not related to dysphonia
No FHS or HR-QoL
Questionnaire not in English
Eligibility
Number of questionnaires assessed for eligibility: 67
Number of questionnaires excluded: 52
Less than 1/2 of the items FHS and/ or HR-QoL related
Not self-reported questionnaire
Not adult population
Singing population
Generic QoL questionnaire
Questionnaire not in English
Included
Total number of questionnaires included: 15
Search 2: to retrieve publications on the psychometric properties of the retrieved questionnaires
Identification
Number of records identified through Embase: 334
Number of records identified through Pubmed: 731
Number of records after duplicates removed: 937
Screening
Number of records screened: 937
Number of records excluded: 816
Not related to any of the 14 included questionnaires
Not psychometrics reported
Article not in English
Unable to be sourced
Eligibility
Number of full-text articles assessed for eligibility: 121
Number of full-text articles excluded: 76
Not psychometrics reported
Report on responsiveness or cross cultural validity
Included
Number of articles added through reference lists: 2
Number of manuals located: 0
Total number of articles included: 47
Total number of assessments included: 15
Total number of articles included: 47

Table 5. *Overview of dysphonia related FHS and HR-QoL questionnaires: reasons for exclusion*

No.	**Questionnaire** (alphabetical order)	**Acronym**	**Exclusion**
1	Behavior Assessment Battery [39]	BAB	Not a measure of FHS and/or HR-QoL in dysphonia
2	Category Ratio scale [40]	CR-10	Not a comprehensive measure
3	BORG CR-10 adapted for Vocal Effort Ratings [41]	N/A	Not a comprehensive measure
4	Children's Voice Handicap Index [42]	CVHI-10	Different target population
5	Classical Singing Handicap Index [43]	CSHI	Not developed in English
6	Communicative Participation Item Bank [44]	CPIB	Not a measure of FHS and/or HR-QoL in dysphonia
7	Condição de Produção Vocal – Professor (Vocal Production Condition – Teacher) [45]	CVV-P (CPV-T)	Not developed in English
8	Disease-specific Self-Efficacy Spasmodic Dysphonia Scale [46]	SE-SD	Not a measure of FHS and/or HR-QoL in dysphonia
9	Dysphonia Risk Screening Protocol [47]	DRSP	Not a measure of FHS and/or HR-QoL in dysphonia
10	Dysphonia Severity Index [48]	DSI	Not a measure of FHS and/or HR-QoL in dysphonia
11	Glasgow Benefit Index [49]	GBI	Not a measure of FHS and/or HR-QoL in dysphonia
12	Glasgow Children's Benefit Inventory [50]	GCBI	Not a measure of FHS and/or HR-QoL in dysphonia
13	Goldman Voice Use Survey [51]	N/A	Not a measure of FHS and/or HR-QoL in dysphonia
14	Iowa Patient's Voice Index (IPVI) [52]	IPVI	Not a comprehensive measure
15	Iranian Voice Quality of Life Profile [53]	IVQLP	Not developed in English
16	Levels of Speech Usage scale [54]	N/A	Not a measure of FHS and/or HR-QoL in dysphonia
17	Linear Analog Scale of Assessment Voice Quality [55]	LASA-VQ	Not a comprehensive measure
18	Living with Dysarthria [56, 57]; Questionário Vivendo com Disartria (VcD; Spanish versión)	LwD (VcD)	Not a measure of FHS and/or HR-QoL in dysphonia
19	Manchester Short Assessment of Quality of Life [58]	MANSA	Not a measure of FHS and/or HR-QoL in dysphonia
20	Modern Singing Handicap Index (Índice de Desvantagem para o Canto Moderno) [59]	MSHI (IDCM)	Not developed in English
21	Occupational Voice Demands Scale [60]	N/A	Not a measure of FHS and/or HR-QoL in dysphonia
22	Pediatric Voice Handicap Index [61]	pVHI	Different target population
23	Pediatric Voice Outcome Survey [62]	Paediatric VOS	Different target population
24	Pediatric Voice-Related Quality-of-Life Survey [63]	PVRQOL	Different target population
25	Quality of Life Questionnaire – Voice Clinic Questionnaire [64]	N/A	Not a comprehensive measure
26	Rastrea-mento de Alterações Vocais em Idosos (Screening for Voice Disorders in Older Adults) [65]	RAVI	Not developed in English

No.	**Questionnaire** (alphabetical order)	Acronym	Exclusion
27	Scleroderma Logopedic Scale [66]	SLS-Voice	Not a self-reported measure
28	Self-rating scale of vocal impairment [67]	SRS	No comprehensive measure
29	Sheehan Disability Scale [68]	SDS	Not a measure of FHS and/or HR-QoL in dysphonia
30	Singers's Vocal Behaviours Questionnaires [69]	Q-SinVocHab	Not developed in English
31	South Manchester University Hospitals NHS Trust: Voice Rating Scale [70]	N/A	Not a comprehensive measure
32	Speech Disability Questionnaire [71]	SDQ	Different target population
33	Stanford Presenteeism Scale 6 [72]	SPS-6	Not a measure of FHS and/or HR-QoL in dysphonia
34	Symptom-specific questionnaire for patients with recurrent papillomatosis [73]	N/A	Not a comprehensive measure
35	Survey on teaching performing [74]	N/A	Not a measure of FHS and/or HR-QoL in dysphonia
36	Thyroidectomy Related Voice Questionnaire [75]	TVQ	Not developed in English
37	Unified Spasmodic Dysphonia Rating Scale [76]	USDRS	Not a self-reported measure
38	University of Rhode Island Change Assessment– Voice [77]	URICA-Voice	Not a measure of FHS and/or HR-QoL in dysphonia
39	Visual Analogue Scale – Voice/ Three Item Outcome Scale [78]	VAS-voice	Not a comprehensive measure
40	Vocal Fatigue Handicap Questionnaire [79]	VFHQ	Not developed in English
41	Vocal Tract Discomfort Scale [80]	VTD	Not a measure of FHS and/or HR-QoL in dysphonia
42	Voice Activity and Participation Profile [81]	VAPP	Not developed in English
43	Voice Disorder Outcome Profile [82]	Voice-DOP	Not developed in English
44	Voice Handicap Index-12 (Stimmstörungsindex) [83]	VHI-12 (SSI)	Not developed in English
45	Voice Handicap Index-13 [84]	VHI-13	Not developed in English
46	Voice Handicap Index-Partner [85]	VHI-P	Not a self-reported measure
47	Voice Handicap Index-Throat [86]	VHI-T	Not developed in English
48	Voice Problems Self-Assessment Scale [87]	VPSS	Not developed in English
49	Voice-Related Statements [72]	VRS	Not a comprehensive measure
50	Voice Self-Efficacy Questionnaire [88]	VSEQ	Not a measure of FHS and/or HR-QoL in dysphonia
51	Voice Outcome Survey [89]	VOS	Not a comprehensive measure
52	Work Productivity And Impairment: Specific Health Problem [90] adapted for hoarseness and voice disorders [91]	WPAI-SHP*	Not a measure of FHS and/or HR-QoL in dysphonia

Notes. No = number; N/A = Not Applicable (no acronym); *Adapted for hoarseness and voice disorders

Table 6. *Studies for the development and validation of questionnaires assessing FHS and/or HR-QoL in dysphonia.*

Questionnaire *(alphabetical order)*	**Reference**	**Purpose of study**	**Study population**	**Age (range [R] and/or Mean [MN] and/or Standard Deviation [SD])**
Evaluation of the Ability to Sing Easily (EASE)	Phyland DJ, Pallant JF, Benninger MS, Thibeault SL, Greenwood KM, Smith JA and Vallance N [20]	To devise, test and validate a scale to assess singer's perceptions of current status of the physical functioning of their singing voice	N = 284: professional music theatre singers (F = 157; M = 127); Age groups: ≤17y (n = 5), 18-20y (n = 12), 21-29y (n = 137), 30-39y (n = 79), 40-49y (n = 41), 50-59y (n = 10)	R = 17-59y, MN = NR, SD = NR
	Phyland DJ, Pallant JF, Thibeault SL, Benninger MS, Vallance N and Smith JA [92]	To use EASE to assess professional music theatre singers' perceptions of current singing voice status and to compare scores across demographic and performance characteristics and to evaluate the construct validity (known-groups validity) of the EASE and its subscales.	N = 284: professional music theatre singers (F = 157; M = 127); Age groups: 17-20y (n = 17), 21-29y (n = 137), 30-39y (n = 79), 40-49y (n = 41), 50-59y (n = 10)	R= 17-59y, MN = NR, SD = NR
Glottal Function Index (GFI)	Bach KK, Belafsky PC, Wasylik K, Postma GN and Koufman JA [21]	To test the reproducibility, specificity and construct validation of the GFI (correlation with VHI)	N = 200; (I) n = 40: patients undergoing laryngoplastic phonosurgery due to glottic insufficiency; (II) n = 40: nodules; (III) n = 40: spasmodic dysphonia (IV) n = 40: granuloma; (V) n = 40: controls (F = 20; M = 20)	(I) R = NR, Med = 49y, SD = NR; (II - IV) R = NR, MN = NR, SD = NR; (V) R = NR, Med = 39y, SD = NR
	Buckmire RA, Bryson PC and Patel MR [93]	To determine the effectiveness of gore-tex medialisation thyroplasty for the management of glottic incompetence in patients with mobile vocal folds and to evaluate the construct validation of GFI (correlation with VRQ & GRBAS)	N = 22: patients with glottis incompetence and mobile vocal folds treated surgically with gore-tex medialisation thyropolasty (F = 12; M = 8)	R = 19-85y, MN = 48 y, SD = NR
Singing Voice Handicap Index (SVHI)	Castelblanco L, Habib M, Stein DJ, de Quadros A, Cohen SM and Noordzij JP [94]	To evaluate the relationship between perceived singing voice health as measured by the SVHI and videostroboscopy findings in healthy professional singers.	N = 47: adult self-reported vocally-healthy professional singers (F = NR; M = NR)	R = 19-62y, MN = 31y, SD = NR
	Cohen SM, Jacobson BH, Garrett CG, Noordzij JP, Stewart MG, Attia A, Ossoff RH and Cleveland TF [22]	To create and validate a singer self-report scale to evaluate the impact of a voice problem on their quality of life	N = 241; (I) n = 112: dysphonic singers (F = 67; M = 45); (II) n = 129: normal singers (F = NR; M = NR)	(I) R = 16-67y, MN = 35.3y, SD = NR; (II) R = NR, MN = NR, SD = NR
Singing Voice Handicap Index-10 (SVHI-10)	Cohen SM, Statham M, Rosen CA and Zullo T [23]	Prospective item analysis of the SVHI and creation and testing of the shortened SVHI and correlation with VHI	N = 297; (I) n= 91: treatment-seeking singers with singing voice problems (F = 62; M = 29); (II) n = 99: non-treatment seeking singers (F = 70; M = 29)	(I) R = 18-33y, MN = 40y, SD = NR; (II) R = 18-80y, MN = 39.5y, SD = NR
Transgender Self-Evaluation Questionnaire (TSEQ)	Hancock AB, Krissinger J and Owen K [95]	To explore relationships between self-perceived quality of life and perceptions of femininity and likability associated with transgender voice	N = 20: male to female transgender individuals	R = 23–63y, MN = 45.8y, SD= 10.6
	Hancock AB [96]	To describe the voice-related quality of life of a diverse group of transgender women using the ICF framework	N = 32: transgender women	R = 19-80y, MN = 43 y, SD = 15 (*Note*: data from N = 81 transgender women, but no specific data for subgroup N = 32)
Transsexual Voice Questionnaire -	Dacakis G, Davies S, Oates JM, Douglas JM	The development and preliminary psychometric evaluation of the TVQ^{MtF}	N = 35: male to female transsexuals	R = 29.8–67.0y, MN = 52.90y, SD = NR

Questionnaire *(alphabetical order)*	**Reference**	**Purpose of study**	**Study population**	**Age (range [R] and/or Mean [MN] and/or Standard Deviation [SD])**
Male to Female (TVQ^{MtF})	and Johnston JR [25]			
Vocal Fatigue Index (VFI)	Nanjundeswaran C, Jacobson BH, Gartner-Schmidt J and Verdolini Abbott K [26]	The development and preliminary psychometric evaluation of the VFI, to help identify individuals with vocal fatigue and characterise their complaints	N = 200 (Stage: Index development and initial testing): (I) patients presenting to voice clinics (F = 127; M = 73) N = 175 (Stage: Validation of VFI): (II) n = 105: voice patients (F = 70; M = 35); (III) n = 70: vocally healthy individuals (F = 49; M = 21)	(I) R = NR, MN = ±51y, SD ≈16-20y; (II) R = NR, MN = 50y, SD = 16; (II) R = NR, MN = 39y, SD = 15
Vocal Performance Questionnaire (VPQ)	Carding PN and Horsley IA [27]	To evaluate the effectiveness of both direct and indirect therapy in the treatment of non-organic dysphonia and survey individual patient responses to these approaches	N = 30: treatment-seeking patients with non-organic dysphonia (F = 23, M = 7); (I) n = 10: direct therapy (F = 8, M = 2); (II) n = 10: indirect therapy (F = 8, M = 2); (III) n = 10: No therapy (F = 7, M = 3)	(I) R = 22-48y, MN = 38.8y, SD = NR; (II) R = 18-76y, MN = 50.1y, SD = NR; (II) R = 18-75y MN = 44.3y, SD = NR
	Deary IJ, Webb A, Mackenzie K, Wilson JA and Carding PN [97]	To evaluate the psychometric properties and clinical usefulness of the VHI-10 and the VPQ in the laryngology office setting	N = 330: dysphonic voice clinic attendees (F = 222, M = 108)	(Males) R = 23-88y, MN = 55.2y, SD = 15.5; (Females) R =17-87y, MN = 48.2y, SD= 17.5
	Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen IN and Wilson JA [98]	To evaluate the reliability and validity of the VPQ, VHI and the VoiS	N = 170: voice clinic patients (F = 127, M = 54)	R = 18-88y, MN= 52 y, SD = NR
Voice Capabilities Questionnaire (VCQ)	Buckley KL, O'Halloran PD and Oates JM [28]	To explore the occupational voice use and vocal health of elite football coaches	N = 12: Australian football coaches (F = 0, M = 12)	R = 32-48y, MN = 39y, SD= 6.7
Voice Disability Coping Questionnaire (VDCQ)	Epstein R, Hirani SP, Stygall J and Newman SP [29]	To explore and validate the VDCQ	N = 80 (F = 52, M = 28): (I) n = 40: voice clinic patients with adductor spasmodic dysphonia (F = 23, M = 17); (II) n = 40: patients with muscle tension dysphonia (F = 29, M = 11)	Total group: R = NR, MN = 45.4y, SD = NR; (I) R = NR, MN = 49.70y, SD = 16.28; (II) R = NR, MN = 41.31y, SD = 19.57
Voice Handicap Index (VHI)	Awan SN, Roy N and Cohen SM [99]	To examine the strength of relationships between acoustic measures, the Cepstral Spectral Index of Dysphonia (CSID) and the total VHI score in voice- disordered and control subjects	N = 332 (F = 216, M = 116): (I) n = 258: voice-disordered patients; (II) n = 74: normophonic control subjects	R = 15–87y, MN = 51.94y, SD= 16.22
	Bach KK, Belafsky PC, Wasylik K, Postma GN and Koufman JA [21]	To test the reproducibility, specificity and construct validation of the GFI (correlation with VHI)	N = 200; (I) n = 40: patients undergoing laryngoplastic phonosurgery due to glottic insufficiency; (II) n = 40: nodules; (III) n = 40: spasmodic dysphonia (IV) n = 40: granuloma; (V) n = 40: controls (F = 20; M = 20)	(I) R = NR, Med = 49y, SD = NR; (II - IV) R = NR, MN = NR, SD = NR; (V) R = NR, Med = 39y, SD = NR
	Elam JC, Ishman SL, Dunbar KB, Clarke JO and Gourin CG [100]	To determine if a relationship exists between depression and VHI-scores in patients with laryngopharyngeal reflux disease	N = 89 treatment–seeking laryngology patients: (I) n = 36 patients with laryngopharyngeal reflux disease (F = 25, M = 11); (II) n = 53 control patients (F = 36, M = 17)	(I) R = 28–77y, MN = 54.2y, SD = NR; (II) R = 19–84y, MN = 48.2y, SD = NR
	Ford Baldner E, Doll E and van Mersbergen MR	To investigate potential relationships between the Borg Category Ratio (CR-10) for vocal effort and auditory-	N = 56: (I) n= 28: participants with voice disorders (F = 17, M = 10, gender undisclosed = 1) (II) n =	(I) R = NR, MN = 50y, SD = NR; (II) R = NR, MN = 27y, SD = NR

Questionnaire *(alphabetical order)*	**Reference**	**Purpose of study**	**Study population**	**Age (range [R] and/or Mean [MN] and/or Standard Deviation [SD])**
	[41]	perceptual ratings, VHI and phonation threshold pressure	28: healthy controls (F = 25, M = 2)	
	Fulljames N and Harris S [101]	To compare correlations between VHI and a self-rating scale of vocal impairment (SRS) with perceptual assessment and acoustic analysis of sustained vowels and connected speech	N = 10: pre and post voice therapy (F = 10, M = 0)	R = 17-76y, MN = 52y, SD = 17.5
	Jacobson BH, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger MS and Newman CW [30]	To develop and validate the VHI	N = 65 (Stage: Scale development): (I) voice clinic patients (F = 40, M = 25) N = 63 (Stage: Test-retest reliability & Relationship of VHI to voice disorder severity): (II) voice clinic patients (F = 38, M = 25)	(I) R = NR, MN = 52.3y, SD = 16.28; (II) R = NR, MN = 49y, SD = 18
	Kazi R, De Cordova J, Singh A, Venkitaraman R, Nutting CM, Clarke P, Rhys-Evans P and Harrington KJ [102]	To investigate the effect of voice impairment across the physical, emotional, and functional domains in patients using valved speech following total laryngectomy as measured by the V- RQOL and the VHI	N = 54: patients after total laryngectomy (F = 14, F = 40)	R = 37–84y, Med = 63.4y, SD = NR
	Portone CR, Hapner ER, McGregor L, Otto K and Johns MM, 3rd [103]	To investigate the correlation between the VHI and the V-RQOL, and to test conversion of scores between the two instruments.	N = 140: voice-disordered patient charts	R = NR, MN = NR, SD = NR
	Rosen CA, Lee AS, Osborne J, Zullo T and Murry T [31]	To develop an abbreviated voice handicap assessment instrument and compare it with the VHI	N = 819: voice clinic patients	R = NR, MN = NR, SD = NR
	Stomeo F, Tosin E, Morolli F, Bianchini C, Ciorba A, Pastore A and Pelucchi S [104]	To evaluate voice outcomes, in patients with early glottic cancer who underwent transoral laser cordectomy with objective and subjective (VHI and GIRBAS scale) means.	N = 24 patients with early glottic cancer who underwent transoral laser cordectomy (F = 1, M = 23)	R = 45 to 86y, Med = 66.4y, SD = NR
	Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen IN and Wilson JA [98]	To evaluate the reliability and validity of the VPQ, VHI and the VoiSS	N = 170: voice clinic patients (F = 127, M = 54)	R = 18-88y, MN= 52 y, SD = NR
	Wheeler KM, Collins SP and Sapienza CM [105]	To examine the relationship between the VHI and acoustic measures of voice samples common in clinical practice.	N = 50: voice-disordered participants completed VHI (F = 38, M = 12) (*Note*: n = 17 for acoustic analysis of voice)	R = 19-80y, MN = 49 yrs, SD = NR
	Wilson JA, Webb A, Carding PN, Steen IN, MacKenzie K and Deary IJ [35]	To perform a substantial factor analysis of two measures of voice impairment, the VoiSS and the VHI	N = 319: voice-disordered patients (F = 220, M = 99); (II) n = 144: functional dysphonics (F = 106, M = 38); n = 145: patients with defined pathology (F = 85, M = 60)	(I) R = NR, MN = 53.4y, SD = 16.5; (II) R = NR, MN = 51.6y, SD=17.3
Voice Handicap Index-10 (VHI-10)	Childs LF, Bielinski C, Toles L, Hamilton A, Deane J and Mau T [106]	To determine if a correlation exists between the VHI-10 and the Voice Functional Communication Measure rating in the National Outcomes Measurement System (NOMS)	N = 409: treatment-seeking voice patients (F = 284, M = 125)	R= 9-90y, MN = 53.1y, SD = NR

Questionnaire *(alphabetical order)*	**Reference**	**Purpose of study**	**Study population**	**Age (range [R] and/or Mean [MN] and/or Standard Deviation [SD])**
	Cohen SM, Statham M, Rosen CA and Zullo T [23]	Prospective item analysis of the SVHI and creation and testing of the shortened SVHI and correlation with VHI	N = 297; (I) n= 91: treatment-seeking singers with singing voice problems (F = 62; M = 29); (II) n = 99: non-treatment seeking singers (F = 70; M = 29)	(I) R = 18-33y, MN = 40y, SD = NR; (II) R = 18-80y, MN = 39.5y, SD = NR
	Davis KM, Sandage MJ, Plexico L and Pascoe DD [107]	To describe athlete belief of performance benefit when using voice during force production tasks and to determine if vocalisation during effortful tasks was correlated to perception of voice impairment	N = 378 athletes (F = 257, M = 121)	R=19->70y, MN = NR, SD = NR
	Deary IJ, Webb A, Mackenzie K, Wilson JA and Carding PN [97]	To evaluate the psychometric properties and clinical usefulness of the VHI-10 and the VPQ in the laryngology office setting	N = 330: dysphonic voice clinic attendees (F = 222, M = 108)	(Males) R = 23-88y, MN = 55.2y, SD = 15.5; (Females) R =17-87y, MN = 48.2y, SD= 17.5
	Eadie TL, Lamvik K, Baylor CR, Yorkston KM, Kim J and Amtmann D [108]	To determine how the self-report outcome measure, the Communicative Participation Item Bank (CPIB), relate to disease- and discipline-specific quality of life outcomes in a head and neck cancer population.	N = 195: individuals treated for head and neck cancer (F = 76, M = 119)	M=61yrs, SD=12.3, R= 24-86
	Gillespie AI, Gooding W, Rosen C and Gartner-Schmidt J [109]	To correlate change in VHI-10 scores with corresponding voice laboratory measures across five voice disorders.	N = 150: voice clinic patients with primary diagnosis of vocal fold lesions, primary muscle tension dysphonia-1, atrophy, unilateral vocal fold paralysis and scar.	R = NR (>18y), MN = NR, SD = NR
	Hu A, Hillel A and Meyer T [110]	To identify and quantify patient perception of perceived hoarseness and impact in spasmodic dysphonia patients.	N = 139: patients with adductor spasmodic dysphonia (F = 106, M = 33)	R = NR, MN = 59.6y, SD= 13.7
	Kupfer RA, Cadalli Tatar E, Barry JO, Allen CT and Merati AL [111]	To determine whether the Derkay score for quantifying papillomatosis positively correlates with the VHI-10	N = 46: recurrent respiratory papillomatosis patients (F = 10, M = 36)	R = NR, MN = NR, SD = NR
	Nichols B, Bock JM and Blumin JH [112]	To identify prevalence of dysphonia in nursing home residents and assisted living residents and search for correlations between VHI-10 and indices of frailty.	N = 119 people in residential care (assisted care and nursing homes); Age groups: 65-75 (n = 35), 76-85 (n = 16), >85 (n = 68)	R= 65->85y, MN=NR, SD = NR
	Romak JJ, Orbelo DM, Maragos NE and Ekbom DC [113]	To examine the correlation between two voice-specific patient-reported outcome measures: VHI-10 and V-RQOL	N = 804 voice clinic patients (F = 478, M = 326)	R = NR, MN = 58.6y, SD=16.3
	Rosen CA, Lee AS, Osborne J, Zullo T and Murry T [31]	To develop an abbreviated voice handicap assessment instrument and compare it with the VHI	N = 819: voice clinic patients	R = NR, MN = NR, SD = NR
	Stachler RJ, Schultz LR, Nerenz D and Yaremchuk KL [114]	to evaluate the Patient-Reported Outcomes Measure Information System (PROMIS) in a head and neck cancer patient cohort by assessing the associations with other measures	N = 39: head and neck patients (F = 10, M = 29)	R = 42-91y, MN = 58.5y, SD=7.7
	Willis J, Michael DD, Boyer H and Misono S [115]	To assess the prevalence and severity of dysphonia in patients with cystic fibrosis sinusitis according to patient-reported outcome measures and auditory-perceptual evaluation	N = 37: participants; (I) n = 17 patients with CF sinusitis; (II) n = 10 patients with non-CF sinusitis; (II) n = 10 healthy individuals	(I) R =18-58y, MN = 30.4y, SD = NR; (II) R = 22-62y, MN = 41.9y, SD = NR; (II) R = 28-37y, MN = 31.3y, SD = NR

Questionnaire *(alphabetical order)*	**Reference**	**Purpose of study**	**Study population**	**Age (range [R] and/or Mean [MN] and/or Standard Deviation [SD])**
Voice Rating Scale (VRS)	Jones SM, Carding PN and Drinnan MJ [70]	To explore whether severity and/or consistency of dysphonia are linked to voice-related quality of life	N = 60: voice clinic patients (F = 36, M = 24)	(Males) R 24–83y, MN = 61.8y, SD = NR; (Females) R = 26–86y, MN = 54.2y, SD = NR
	Wingate JM, Brown WS, Shrivastav R, Davenport P and Sapienza CM [32]	To examine treatment outcomes of two specific rehabilitation programs for a group of professional voice users	N = 18: dysphonic treatment-seeking professional voice-users (F = 10, M = 8): (I) n = 9: benign vocal fold lesions; (II) n = 9: symptoms of throat pain or vocal fatigue	Total group: (Males) R= 27-59y, MN = 46y, SD = NR; Total group: (Females) R = 25-59y, MN = 39y, SD = NR
Voice-Related Quality of Life (V-RQOL)	Bornbaum CC, Day AM and Doyle PC [116]	To evaluate the construct validity of the V-RQOL measure in a sample of individuals who have undergone total laryngectomy	N =109: laryngectomised persons	R = 34-88y, MN = 63.38y, SD = NR
	Buckmire RA, Bryson PC and Patel MR [93]	To determine the effectiveness of gore-tex medialisation thyroplasty for the management of glottic incompetence in patients with mobile vocal folds and to evaluate the construct validation of GFI (correlation with VRQ & GRBAS)	N = 22: patients with glottis incompetence and mobile vocal folds treated surgically with gore-tex medialisation thyropolasty (F = 12; M = 8)	R = 19-85y, MN = 48 y, SD = NR
	Hogikyan ND and Sethuraman G [33]	To develop and validate a clinically useful instrument for measuring voice-related quality of life	N = 31: (I) n = 109: voice patients; (II) n = 22: non-voice patients 22 non-voice patients.	(I) R = 19-85y, MN = 51.2y, SD = NR; (II) R = 19-84, MN = 49.9y, SD = NR
	Karnell MP, Melton SD, Childes JM, Coleman TC, Dailey SA and Hoffman HT [52]	To test reliability and construct validity of the IPVI with GRBAS, CAPE-V and V-RQOL	N = 103: treatment-seeking voice disordered patients (F = 61, M = 42); Age groups: 17-29y (n = 15), 30-39y (n = 12), 40-49y (n = 25), 50-59y (n = 16), 60-69y (n = 12), 70-79y (n = 14), >80y (n = 9)	R=17->80, MN = NR, SD = NR
	Kazi R, De Cordova J, Singh A, Venkitaraman R, Nutting CM, Clarke P, Rhys-Evans P and Harrington KJ [102]	To investigate the effect of voice impairment across the physical, emotional, and functional domains in patients using valved speech following total laryngectomy as measured by the V- RQOL and the VHI	N = 54: patients after total laryngectomy (F = 14, F = 40)	R=37–84y, Med = 63.4y, SD = NR
	Kupfer RA, Hogikyan EM and Hogikyan ND [117]	To establish a large, heterogenous V-RQOL normative database.	N = 253: people who did not perceive themselves to have a voice problem (F = 170, M = 83)	R= 18-94y, MN = 40.9y, SD = NR
	Murry T, Medrado R, Hogikyan ND and Aviv JE [118]	To determine the relationship between the patient's perception of voice related quality of life using the V-RQOL and the clinician's perception of voice severity using the GRBAS	N = 95: (I) n = 50 patients with a complaint of a voice disorder (F = 29, M = 21); (II) n = 45 (F = 26, M = 19): control group without voice complaints	(I) R = 22-90y, M = 59.7y; (II) R = 25-90y, MN = 53.4y
	Portone CR, Hapner ER, McGregor L, Otto K and Johns MM, 3rd [103]	To investigate the correlation between the VHI and the V-RQOL, and to test conversion of scores between the two instruments.	N = 140: voice—disordered patient charts	R = NR, MN = NR, SD = NR
	Romak JJ, Orbelo DM, Maragos NE and Ekbom DC [113]	To examine the correlation between two voice-specific patient-reported outcome measures: VHI-10 and V-RQOL	N = 804 voice clinic patients (F = 478, M = 326)	R = NR, MN = 58.6y, SD=16.3

Questionnaire *(alphabetical order)*	**Reference**	**Purpose of study**	**Study population**	**Age (range [R] and/or Mean [MN] and/or Standard Deviation [SD])**
	Tanner K, Pierce JL, Merrill RM, Miller KL, Kendall KA and Roy N [119]	To examine quality of life burden of voice disorders in Sjögren's syndrome	N = 101: patients with Sjögren's syndrome	R = 20-93y, MN = 59.4y, SD = 14.1
Voice Symptom Scale (VoiSS)	Deary IJ, Wilson JA, Carding PN and MacKenzie K [34]	To devise and validate a patient- derived inventory of voice symptoms for use as a sensitive assessment tool of baseline pathology and response to change in adult dysphonia clinics	N = 180 (Final stage): voice-disordered treatment-seeking patients (F = 117, M = 63)	(Females) R = NR, MN = 53.4y, SD = 16.0; (Males) R = NR, MN = 55.4y, SD=14.0
	Jones SM, Carding PN and Drinnan MJ [70]	To explore whether severity and/or consistency of dysphonia are linked to voice-related quality of life.	N = 60: voice clinic patients (F = 36, M = 24)	(males) R 24–83y, MN = 61.8y, SD = NR; (females) R = 26–86y, MN = 54.2y, SD = NR
	Montgomery J, Hendry J, Wilson JA, Deary IJ and MacKenzie K [120]	To evaluate diagnostic performance of the emotional domain of the VoiSS questionnaire compared with the Hospital Anxiety and Depression Scale (HADS)	N = 177: voice clinic patients (F = 125, M = 52) (*Note*: 177 from 210 consecutive patients)	R= 18–90y, MN = 56.0y, Med = 57.2y, SD = NR
	Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen IN and Wilson JA [98]	To evaluate the reliability and validity of the VPQ, VHI and the VoiSS	N = 170: voice clinic patients (F = 127, M = 54)	R = 18-88y, MN= 52 y, SD = NR
	Wilson JA, Webb A, Carding PN, Steen IN, MacKenzie K and Deary IJ [35]	To perform a substantial factor analysis of two measures of voice impairment, the VoiSS and the VHI	N = 319: voice-disordered patients (F = 220, M = 99); (II) n = 144: functional dysphonics (F = 106, M = 38); n = 145: patients with defined pathology (F = 85, M = 60)	(I) R = NR, MN = 53.4y, SD = 16.5; (II) R = NR, MN = 51.6y, SD=17.3

Note. N = total sample size; n = subgroups; R = range; MN = mean; Med = median; NR = not reported; SD = standard deviation; M = male; F = female

Table 7. *Characteristics of the questionnaires for the assessment of FHS and/or HR-QoL in dysphonia*

Questionnaire *(alphabetical order)*	**Purpose of instrument**	**Published year**	**Scale titles (number of items)**	**Number of scales (total number of items); Range of score**	**Response Options**
Evaluation of the Ability to Sing Easily (EASE) [20]	Self-report singing voice function scale	2014	EASE (22); Subscales Vocal Fatigue (10), Pathological-Risk Indicator (10), Vocal Concern (2)	3 (22); Range: 22-88	4-point ordinal scale (1-4: not at all, mildly, moderately, extremely)
Glottal Function Index (GFI) [21]	Self-report symptom impairment due to glottic insufficiency	2005	Glottic Function Index (4)	1 (4): Range: 0-20	6-point ordinal scale (0-5: no problem to severe problem)
Singing Voice Handicap Index (SVHI) [22]	Voice-related health status questionnaire for singers with voice problems	2007	Singing Voice Handicap Index (36)	1 (36); Range: 0-144	5-point ordinal scale (0-4: never to always)
Singing Voice Handicap Index-10 (SVHI-10) [23]	Shortened version of the Singing Voice Handicap Index (SVHI).	2009	SVHI-10 (10)	1 (10); Range: 0-40	5-point ordinal scale (0-4: never to always)
Transgender Self-Evaluation Questionnaire (TSEQ) [24]	Self-report QoL measure for impact of voice issues tailored to the transgender population	2006	TSEQ (30): Subscales Functional (11), Physical Score (10), Emotional Score (9); (TSEQ Overall (2): separate from TSEQ)	3 (30); Range: 30-150 (TSEQ Overall: NA)	5-point ordinal scale (1-5: never to always); (TSEQ Overall: 5 categorical descriptors)
Transsexual Voice Questionnaire - Male to Female (TVQ^{MtF}) [25]	Self-reported perceptions of MtF transsexuals regarding their voice	2013	TVQ^{MtF} (30)	1 (30); Range: 30-120	4-point ordinal scale (1-4: never, rarely, usually, always).
Vocal Fatigue Index (VFI) [26]	Self-report scale for identifying and characterising the degree and nature of vocal fatigue symptoms	2015	VFI (19): Subscales Tiredness and avoidance of voice (11), Physical discomfort (5), Improvement of symptoms with rest (3)	3 (19); Range: 0-76	5-point ordinal scale (0-4: never, almost never, sometimes, almost always, always)
Vocal Performance Questionnaire (VPQ) [27]	Self-rated survey of patient perceptions of their own vocal performance	1992	VPQ (12)	1 (12); Range 12-60	5-point ordinal scales (a-e: normal voice to severely limited)

Questionnaire *(alphabetical order)*	**Purpose of instrument**	**Published year**	**Scale titles (number of items)**	**Number of scales (total number of items); Range of score**	**Response Options**
Voice Capabilities Questionnaire (VCQ) [28]	Self-rated inventory related to perceived vocal health at work	2015	VCQ (16) (Voice problem (1-2): separate from VCQ)	1 (16); Range: 16-80 (Voice problem: NA)	5-point ordinal scale (1-5: never to always) Voice problem: mixed (binary score: yes/no; if yes, 5-point ordinal scale (1-5): slight to severe)
Voice Disability Coping Questionnaire (VDCQ) [29]	Voice-specific measure to evaluate how patients cope with voice problems.	2009	VDCQ (15): Subscales Social Support (6), Passive coping (4) Avoidance (3), Information- Seeking (2)	4 (15); Range 0-75	6-point ordinal scale (0-5: never to always)
Voice Handicap Index (VHI or VHI-30) [30]	Self-report instrument to quantify the perception of psychosocial and voice-related quality of life impact of a voice problem	1997	VHI (30) Subscales Emotional (10), Functional (10). Physical (10)	3 (30); Range 0-120	5-point ordinal scale (0-4: none to always)
Voice Handicap Index-10 (VHI-10) [31]	Abbreviated voice handicap assessment instrument to quantify the perception of impact of a voice disorder on voice-related quality of life	2004	VHI-10 (10) No subscales, but items are identified as Emotional (2), Functional (5) and Physical (3)	1 (10); Range 0-40	5-point ordinal scale (0-4: never to always)
Voice Rating Scale (VRS) [32]	Self-rating of vocal function in voice-related work activities	2007	VRS (10)	1 (10); Range 0 – 1000?	Visual Analogue scale (mild to moderate to severe)
Voice-Related Quality of Life (V-RQOL) [33]	Self-report scale for evaluation of the perceived impact of a voice disorder on a person's quality of life	1999	Total (10): Subscales Physical Functioning (6), Social-emotional (4)	2 (10); Range 0-100 (domain and total V-RQOL converted raw scores to standardised scores)	5-point ordinal scales (0-5: none/not a problem to problem is as 'bad it can be').
Voice Symptom Scale (VoiSS) [34]	Self-report tool to assess symptoms and psychosocial impact of voice disorders patient- derived items.	2003	Total (30): Subscales Impairment (15), Emotional (8), Physical (7)	3 (10); Range (0-120)	5-point ordinal scales (0-4: never to always)

Table 8. *Methodological quality assessment of studies on psychometric properties of FHS and/or HR-QoL questionnaires in dysphonia.*

Questionnaire *(alphabetical order)*	**Reference**	**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)	Phyland DJ, Pallant JF, Benninger MS, Thibeault SL, Greenwood KM, Smith JA and Vallance N [20]	Excellent (85.7%)	NR	NR	Good (71.4%)	Excellent (75.0%)	NR
	Phyland DJ, Pallant JF, Thibeault SL, Benninger MS, Vallance N and Smith JA [92]	Excellent (81.0%)	NR	NR	NR	NR	Divergent (Gender) Fair (47.1%) Divergent (Age) Fair (47.1%) Divergent (Self-reported voice problem) Fair (41.2%) Divergent (Currently performing) Fair (41.2%)
Glottal Function Index (GFI)	Bach KK, Belafsky PC, Wasylik K, Postma GN and Koufman JA [21]	NR	*Intra-rater* Good (51.7%)	NR	NR	NR	*Convergent (VHI)* Fair (30.4%)
	Buckmire RA, Bryson PC and Patel MR [93]	NR	NR	NR	NR	NR	*Convergent (V-RQOL)* Fair (39.1%)
Singing Voice Handicap Index (SVHI)	Castelblanco L, Habib M, Stein DJ, de Quadros A, Cohen SM and Noordzij JP [94]	NR	NR	NR	NR	NR	Convergent (Videostroboscopy) Fair (47.8%)
	Cohen SM, Jacobson BH, Garrett CG, Noordzij JP, Stewart MG, Attia A, Ossoff RH and Cleveland TF [22]	Good (66.7%)	Good (62.1%)	NR	Good (50.0%)	Fair (50.0%)	Convergent (VAS) Fair (43.5%) Divergent (self-reported diagnosis) Fair (35.3%)
Singing Voice Handicap Index-10 (SVHI-10)	Cohen SM, Statham M, Rosen CA and Zullo T [23]	Good (60.0%)	Good (58.6%)	NR	Fair (50.0%)	NR	Convergent (VHI-10) Fair (34.8%) Convergent (Self-reported voice disorder) Fair (35.3%)
Transgender Self-Evaluation Questionnaire (TSEQ)	Hancock AB, Krissinger J and Owen K [95]	NR	NR	NR	NR	NR	*Convergent (Femininity Self-rate)* Fair (43.5%) *Convergent (Likeability Self-rate)* Fair (43.5%) *Convergent (Femininity Listener-rate)* Fair (43.5%) *Convergent (Likeability Listener-rate)* Fair (43.5%)
	Hancock AB [96]	NR	NR	NR	NR	NR	*Convergent (VHI)* Good (52.2%)
Transsexual Voice Questionnaire - Male to Female (TVQ^{MtF})	Dacakis G, Davies S, Oates JM, Douglas JM and Johnston JR [25]	Good (52.0%)	*Test-retest* Excellent (86.2%)	NR	NR	NR	NR

Questionnaire *(alphabetical order)*	**Reference**	**Measurement property: methodological quality per study**					
		Internal Consistency	**Reliability**	**Measurement Error**	**Content Validity**	**Structural Validity**	**Hypothesis Testing**
Vocal Fatigue Index (VFI)	Nanjundeswaran C, Jacobson BH, Gartner-Schmidt J and Verdolini Abbott K [26]	Excellent (85.7%)	*Test-retest* Good (72.4%)	NR	Excellent (78.6%)	Good (66.7%)	*Divergent* Good (58.8%)
Vocal Performance Questionnaire (VPQ)	Carding PN and Horsley IA [27]	NR	NR	NR	NR	NR	*Convergent (Buffalo III)* Good (56.5%)
	Deary IJ, Webb A, Mackenzie K, Wilson JA and Carding PN [97]	Fair (36.0%)	NR	NR	NR	Fair (44.4%)	*Convergent (VHI-10)* Fair (43.5%)
	Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen IN and Wilson JA [98]	Good (63.9%)	*Intra-rater* Excellent (75.9%)	NR	NR	NR	*Convergent (VHI)* Good (52.2%) *Convergent (VoiSS)* Good (52.2%) *Convergent (GRBAS)* Good (52.2%)
Voice Capabilities Questionnaire (VCQ)	Buckley KL, O'Halloran PD and Oates JM [28]	Fair (40.0%)	NR	NR	NR	NR	NR
Voice Disability Coping Questionnaire (VDCQ)	Epstein R, Hirani SP, Stygall J and Newman SP [29]	Good (56.0%)	NR	NR	Fair (40%)	Fair (38.9%)	*Convergent (VDQ)* Fair (30.4%) *Convergent (HLOC)* Fair (30.4%) *Convergent (BDI)* Fair (30.4%) *Convergent (STAI)* Fair (30.4%) *Convergent (RSES)* Fair (30.4%)
Voice Handicap Index (VHI)	Awan SN, Roy N and Cohen SM [99]	NR	NR	NR	NR	NR	*Convergent (CAPE-V)* Good (73.9%) *Divergent (Gender)* Good (64.7%)
	Bach KK, Belafsky PC, Wasylik K, Postma GN and Koufman JA [21]	NR	NR	NR	NR	NR	*Convergent (GFI)* Fair (30.4%)
	Elam JC, Ishman SL, Dunbar KB, Clarke JO and Gourin CG [100]	NR	NR	NR	NR	NR	*Convergent (Depression)* Good (52.2%) *Divergent (Reflux)* Good (52.9%)
	Ford Baldner E, Doll E and van Mersbergen MR [41]	NR	NR	NR	NR	NR	*Convergent (Borg)* Good (56.5%) *Convergent (CAPE)*

Questionnaire *(alphabetical order)*	**Reference**	**Measurement property: methodological quality per study**					
		Internal Consistency	**Reliability**	**Measurement Error**	**Content Validity**	**Structural Validity**	**Hypothesis Testing**
							Good (60.9%) *Divergent (Voice disorder)* Fair (47.1%)
	Fulljames N and Harris S [101]	NR	NR	NR	NR	NR	*Divergent (baseline GRBAS)* Good (70.6%) *Divergent (baseline SRS)* Good (70.6%) *Divergent (baseline Acoustics)* Good (70.6%) *Divergent (delta change GRBAS)* Good (70.6%) *Divergent (delta change SRS)* Good (70.6%) *Divergent (delta change Acoustics)* Good (70.6%)
	Jacobson BH, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger MS and Newman CW [30]	Fair (44.0%)	*Intra-rater* Fair (44.8%)	NR	Fair (40.0%)	NR	*Convergent (Voice Disorder Severity)* Good (34.8%)
	Kazi R, De Cordova J, Singh A, Venkitaraman R, Nutting CM, Clarke P, Rhys-Evans P and Harrington KJ [102]	NR	NR	NR	NR	NR	*Convergent (V-RQOL)* Good (70.0%)
	Portone CR, Hapner ER, McGregor L, Otto K and Johns MM, 3rd [103]	NR	NR	NR	NR	NR	*Convergent (V-RQOL)* Good (56.5%) Convergent (VR-QOL diagnosis) Good (52.2%)
	Rosen CA, Lee AS, Osborne J, Zullo T and Murry T [31]	NR	NR	NR	NR	NR	*Convergent (VHI-10)* Excellent (82.5%)
	Stomeo F, Tosin E, Morolli F, Bianchini C, Ciorba A, Pastore A and Pelucchi S [104]	NR	NR	NR	NR	NR	*Convergent (GIRBAS)* Good (52.2%) *Convergent (Accoustic)* Good (52.2%) *Convergent (Yanagihara classification)* Good (52.2%) *Convergent (Max Phonation Time)* Good (52.2%) *Convergent (Laryng Videostroboscopy)* Good (52.2%)
	Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen IN and Wilson JA	Good (63.9%)	*Intra-rater* Good (75.9%)	NR	NR	NR	*Convergent (VPQ)* Good (52.2%) *Convergent (VoiSS)*

Questionnaire *(alphabetical order)*	**Reference**	**Measurement property: methodological quality per study**					
		Internal Consistency	**Reliability**	**Measurement Error**	**Content Validity**	**Structural Validity**	**Hypothesis Testing**
	[98]						Good (52.2%) *Convergent (GRBAS)* Good (52.2%)
	Wheeler KM, Collins SP and Sapienza CM [105]	NR	NR	NR	NR	NR	*Convergent (Acoustic)* Fair (43.6%)
	Wilson JA, Webb A, Carding PN, Steen IN, MacKenzie K and Deary IJ [35]	Good (60.0%)	NR	NR	NR	Good (58.3%)	*Divergent* Poor (17.6%)
Voice Handicap Index-10 (VHI-10)	Childs LF, Bielinski C, Toles L, Hamilton A, Deane J and Mau T [106]	NR	NR	NR	NR	NR	*Convergent* Good (73.9%) *Divergent (Gender)* Good (58.8%) *Divergent (Diagnosis)* Good (58.8%) *Divergent (Smoking*) Good (70.59%) *Divergent (Psychiatric)* Good (70.59%)
	Cohen SM, Statham M, Rosen CA and Zullo T [23]	NR	NR	NR	NR	NR	Convergent (SVHI-10) Fair (34.8%)
	Davis KM, Sandage MJ, Plexico L and Pascoe DD [107]	NR	NR	NR	NR	NR	*Convergent (PPE)* Good (52.2%) *Divergent (Athlete)* Good (52.9%) *Divergent (Use of voice)* Good 58.8%)
	Deary IJ, Webb A, Mackenzie K, Wilson JA and Carding PN [97]	Fair (48.0%)	NR	NR	NR	Fair (44.4%)	*Convergent (VPQ)* Fair (43.5%)
	Eadie TL, Lamvik K, Baylor CR, Yorkston KM, Kim J and Amtmann D [108]	NR	NR	NR	NR	NR	Good (69.6%)
	Gillespie AI, Gooding W, Rosen C and Gartner-Schmidt J [109]	NR	NR	NR	NR	NR	*Divergent (baseline Cape-V)* Excellent (88.2%) *Divergent (baseline Acoustic)* Excellent (88.2%) *Divergent (baseline Aerodynamic)* Excellent (88.2%) *Divergent (delta change Cape-V)* Excellent (88.2%) *Divergent (delta change Acoustic)* Excellent (88.2%) *Divergent (delta change Aerodynamic)*

Questionnaire *(alphabetical order)*	Reference	Measurement property: methodological quality per study					
		Internal Consistency	**Reliability**	**Measurement Error**	**Content Validity**	**Structural Validity**	**Hypothesis Testing**
							Excellent (88.2%)
	Hu A, Hillel A and Meyer T [110]	NR	NR	NR	NR	NR	*Convergent (CAPE-V)* Good (65.2%) *Convergent (HADS Anxiety*) Good (65.2%) *Convergent (HADS Depression)* Good (65.2%) *Divergent (Gender)* Good (52.9%) *Divergent (Age)* Good (52.9%) *Divergent (Time of diagnosis)* Good (58.8%)
	Kupfer RA, Cadalli Tatar E, Barry JO, Allen CT and Merati AL [111]	NR	NR	NR	NR	NR	*Convergent (Derkay total sample)* Fair (47.8%) *Convergent (Excluding juvenile)* Fair (47.8%) *Convergent (Low prior surgery)* Fair (43.5%) *Convergent (High prior surgery)* Fair (47.83%)
	Nichols B, Bock JM and Blumin JH [112]	NR	NR	NR	NR	NR	*Convergent (VES13)* Good (60.9%) *Divergent (Frailty)* Good (58.8%) *Divergent (Nursing home)* Good (58.8%)
	Romak JJ, Orbelo DM, Maragos NE and Ekbom DC [113]	NR	NR	NR	NR	NR	*Convergent (V-RQOL)* Good (68.8%)
	Rosen CA, Lee AS, Osborne J, Zullo T and Murry T [31]	NR	NR	NR	Excellent (85.7%)	NR	*Convergent (VHI)* Excellent (82.5%)
	Stachler RJ, Schultz LR, Nerenz D and Yaremchuk KL [114]	NR	NR	NR	NR	NR	*Convergent (PROMIS)* Good (52.2%)
	Willis J, Michael DD, Boyer H and Misono S [115]	NR	NR	NR	NR	NR	*Divergent (CF diagnosis)* Fair (41.2%)
Voice Rating Scale (VRS)	Jones SM, Carding PN and Drinnan MJ [70]	NR	NR	NR	NR	NR	*Convergent (VoiSS)* Fair (47.8%)
	Wingate JM, Brown WS, Shrivastav R, Davenport P and Sapienza CM [32]	NR	*Intra-rater* Fair (31.0%) *Inter-rater*	NR	NR	NR	*Convergent (VHI)* Good (52.2%)

Questionnaire *(alphabetical order)*	**Reference**	**Measurement property: methodological quality per study**					
		Internal Consistency	**Reliability**	**Measurement Error**	**Content Validity**	**Structural Validity**	**Hypothesis Testing**
			Fair (41.4%)				
Voice-Related Quality of Life (V-RQOL)	Bornbaum CC, Day AM and Doyle PC [116]	NR	NR	NR	NR	Good (55.6%)	NR
	Buckmire RA, Bryson PC and Patel MR [93]	NR	NR	NR	NR	NR	*Convergent (GFI)* Fair (39.1%)
	Hogikyan ND and Sethuraman G [33]	Fair (32.0%)	*Intra-rater* Fair (44.8%)	NR	NR	NR	*Convergent (SF-36)* Fair (26.10%) *Divergent (self-reported voice problems*) Good (52.9)
	Karnell MP, Melton SD, Childes JM, Coleman TC, Dailey SA and Hoffman HT [52]	NR	NR	NR	NR	NR	*Convergent (IPVI)* Good (56.5%) *Convergent (GRBAS)* Good (56.5%) *Convergent (CAPE-V)* Good (56.5%)
	Kazi R, De Cordova J, Singh A, Venkitaraman R, Nutting CM, Clarke P, Rhys-Evans P and Harrington KJ [102]	NR	NR	NR	NR	NR	*Convergent (VHI)* Good (70.0%)
	Kupfer RA, Hogikyan EM and Hogikyan ND [117]	NR	NR	NR	NR	NR	*Convergent (Self-rate)* Good (52.2%) *(Age*)* Good (52.9%) *(Gender*)* Good (52.9%) *(Employment*)* Good (52.9%) *[No expected direction of correlations or mean differences included in hypothesis]
	Murry T, Medrado R, Hogikyan ND and Aviv JE [118]	NR	NR	NR	NR	NR	*Convergent (GRBAS Total; Subgroups per gender, age, voice disorder)* Good (65.2%) *Divergent (GRBAS Control*)* Good (60.9%)
	Portone CR, Hapner ER, McGregor L, Otto K and Johns MM, 3rd [103]	NR	NR	NR	NR	NR	*Convergent (VHI Total)* Good (56.5%) *Convergent (VHI Diagnosis)* Good (52.2%)
	Romak JJ, Orbelo DM, Maragos NE and Ekbom DC [113]	NR	NR	NR	NR	NR	*Convergent (VHI-10)* Good (68.8%)

Questionnaire *(alphabetical order)*	**Reference**	**Measurement property: methodological quality per study**					
		Internal Consistency	**Reliability**	**Measurement Error**	**Content Validity**	**Structural Validity**	**Hypothesis Testing**
	Tanner K, Pierce JL, Merrill RM, Miller KL, Kendall KA and Roy N [119]	NR	NR	NR	NR	NR	*Convergent (SF-36)* Good (56.5%) *Convergent (SSI-Sicca symptoms)* Good (56.5%) *Convergent (ESSPRI-Sicca symptoms)* Good (56.5%) *Divergent (Self-reported Voice disorder)* Good (58.8%)
Voice Symptom Scale (VoiSS)	Deary IJ, Wilson JA, Carding PN and MacKenzie K [34]	NR	NR	NR	Excellent (90%)	NR	NR
	Jones SM, Carding PN and Drinnan MJ [70]	NR	NR	NR	NR	NR	*Convergent (GRBAS)* Good (52.2%) *Convergent (VRS)* Fair (47.8%)
	Montgomery J, Hendry J, Wilson JA, Deary IJ and MacKenzie K [120]	NR	NR	NR	NR	NR	*Convergent (HADS Anxiety)* Good (52.2%) *Convergent (HADS Depression)* Good (52.2%)
	Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen IN and Wilson JA [98]	Good (63.9%)	*Intra-rater* Excellent (75.9%)	NR	NR	NR	*Convergent (VPQ)* Good (52.2%) *Convergent (VHI)* Good (52.2%) *Convergent (GRBAS)* Good 52.2%)
	Wilson JA, Webb A, Carding PN, Steen IN, MacKenzie K and Deary IJ [35]	Good (60.0%)	NR	NR	NR	Good (58.3%)	*Divergent* Poor (17.6%)

Notes. Responsiveness and cross-cultural validity were out of the scope of this review; Criterion validity could not be assessed due to the lack of a 'gold standard' measure in the field of FHS and/or HR-QoL in dysphonia; The methodological quality was determined using the COSMIN rating per item [8]: excellent, good, fair, and poor. The overall methodological quality per study was presented as a percentage of the ratings [16, 17]: Poor = 0-25%, Fair = 25.1% -50.0%, good = 50.1%-75%, Excellent = 75.1%-100.0%; NR = Not Reported.

Table 9. *Quality of measurement properties per study*.

Questionnaire *(alphabetical order)*	**Reference**	**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)	Phyland DJ, Pallant JF, Benninger MS, Thibeault SL, Greenwood KM, Smith JA and Vallance N [20]	**+**	NR	NR	**?**	**+**	NR
	Phyland DJ, Pallant JF, Thibeault SL, Benninger MS, Vallance N and Smith JA [92]	**+**	NR	NR	NR	NR	?
Glottal Function Index (GFI)	Bach KK, Belafsky PC, Wasylik K, Postma GN and Koufman JA [21]	NR	?	NR	NR	NR	+
	Buckmire RA, Bryson PC and Patel MR [93]	NR	NR	NR	NR	NR	+
Transgender Self-Evaluation Questionnaire (TSEQ)	Hancock AB, Krissinger J and Owen K [95]	NR	NR	NR	NR	NR	?
	Hancock AB [96]	NR	NR	NR	NR	NR	+
Singing Voice Handicap Index (SVHI)	Castelblanco L, Habib M, Stein DJ, de Quadros A, Cohen SM and Noordzij JP [94]	NR	NR	NR	NR	NR	?
	Cohen SM, Jacobson BH, Garrett CG, Noordzij JP, Stewart MG, Attia A, Ossoff RH and Cleveland TF [22]	-	?	NR	?	+	+
Singing Voice Handicap Index-10 (SVHI-10)	Cohen SM, Statham M, Rosen CA and Zullo T [23]	+	?	NR	-	NR	+
Transsexual Voice Questionnaire - Male to Female (TVQ^{MtF})	Dacakis G, Davies S, Oates JM, Douglas JM and Johnston JR [25]	?	+	NR	NR	NR	NR
Vocal Fatigue Index (VFI)	Nanjundeswaran C, Jacobson BH, Gartner-Schmidt J and Verdolini Abbott K [26]	+	?	NR	+	-	+
Vocal Performance Questionnaire (VPQ)	Carding PN and Horsley IA [27]win	NR	NR	NR	NR	NR	?
	Deary IJ, Webb A, Mackenzie K, Wilson JA and Carding PN [97]	+	NR	NR	NR	+	+
	Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen IN and Wilson JA [98]	?	+	NR	NR	NR	?
Voice Capabilities Questionnaire (VCQ)	Buckley KL, O'Halloran PD and Oates JM [28]	?	NR	NR	NR	NR	NR
Voice Disability Coping Questionnaire (VDCQ)	Epstein R, Hirani SP, Stygall J and Newman SP [29]	?	NR	NR	+	+	?
Voice Handicap Index (VHI)	Awan SN, Roy N and Cohen SM [99]	NR	NR	NR	NR	NR	?
	Bach KK, Belafsky PC, Wasylik K, Postma GN and Koufman JA [21]	NR	NR	NR	NR	NR	+
	Cohen SM, Statham M, Rosen CA and Zullo T [23]	+	?	NR	-	NR	+

Questionnaire *(alphabetical order)*	Reference	Measurement property: methodological quality per study					
		Internal Consistency	**Reliability**	**Measurement Error**	**Content Validity**	**Structural Validity**	**Hypothesis Testing**
	Elam JC, Ishman SL, Dunbar KB, Clarke JO and Gourin CG [100]	NR	NR	NR	NR	NR	?
	Ford Baldner E, Doll E and van Mersbergen MR [41]	NR	NR	NR	NR	NR	+
	Fulljames N and Harris S [101]	NR	NR	NR	NR	NR	+
	Hancock AB [96]	NR	NR	NR	NR	NR	+
	Jacobson BH, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger MS and Newman CW [30]	?	+	NR	?	NR	+
	Kazi R, De Cordova J, Singh A, Venkitaraman R, Nutting CM, Clarke P, Rhys-Evans P and Harrington KJ [102]	NR	NR	NR	NR	NR	?
	Portone CR, Hapner ER, McGregor L, Otto K and Johns MM, 3rd [103]	NR	NR	NR	NR	NR	+
	Rosen CA, Lee AS, Osborne J, Zullo T and Murry T [31]	NR	NR	NR	NR	NR	+
	Stomeo F, Tosin E, Morolli F, Bianchini C, Ciorba A, Pastore A and Pelucchi S [104]	NR	NR	NR	NR	NR	?
	Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen IN and Wilson JA [98]	?	+	NR	NR	NR	?
	Wheeler KM, Collins SP and Sapienza CM [105]	NR	NR	NR	NR	NR	?
	Wilson JA, Webb A, Carding PN, Steen IN, MacKenzie K and Deary IJ [35]	+	NR	NR	NR	+	NE
Voice Handicap Index-10 (VHI-10)	Childs LF, Bielinski C, Toles L, Hamilton A, Deane J and Mau T [106]	NR	NR	NR	NR	NR	?
	Cohen SM, Statham M, Rosen CA and Zullo T [23]	NR	NR	NR	NR	NR	+
	Davis KM, Sandage MJ, Plexico L and Pascoe DD [107]	NR	NR	NR	NR	NR	?
	Deary IJ, Webb A, Mackenzie K, Wilson JA and Carding PN [97]	+	NR	NR	NR	+	+
	Eadie TL, Lamvik K, Baylor CR, Yorkston KM, Kim J and Amtmann D [108]	NR	NR	NR	NR	NR	?
	Gillespie AI, Gooding W, Rosen C and Gartner-Schmidt J [109]	NR	NR	NR	NR	NR	?
	Hu A, Hillel A and Meyer T [110]	NR	NR	NR	NR	NR	?
	Kupfer RA, Cadalli Tatar E, Barry JO, Allen CT and Merati AL [111]	NR	NR	NR	NR	NR	?
	Nichols B, Bock JM and Blumin JH [112]	NR	NR	NR	NR	NR	?
	Romak JJ, Orbelo DM, Maragos NE and Ekbom DC [113]	NR	NR	NR	NR	NR	+

Questionnaire *(alphabetical order)*	**Reference**	**Measurement property: methodological quality per study**					
		Internal Consistency	**Reliability**	**Measurement Error**	**Content Validity**	**Structural Validity**	**Hypothesis Testing**
	Rosen CA, Lee AS, Osborne J, Zullo T and Murry T [31]	NR	NR	NR	+	NR	+
	Stachler RJ, Schultz LR, Nerenz D and Yaremchuk KL [114]	NR	NR	NR	NR	NR	?
	Willis J, Michael DD, Boyer H and Misono S [115]	NR	NR	NR	NR	NR	-
Voice Rating Scale (VRS)	Jones SM, Carding PN and Drinnan MJ [70]	NR	NR	NR	NR	NR	?
	Wingate JM, Brown WS, Shrivastav R, Davenport P and Sapienza CM [32]	NR	?	NR	NR	NR	?
Voice-Related Quality of Life (V-RQOL)	Bornbaum CC, Day AM and Doyle PC [116]	NR	NR	NR	NR	+	NR
	Buckmire RA, Bryson PC and Patel MR [93]	NR	NR	NR	NR	NR	+
	Hogikyan ND and Sethuraman G [33]	?	+	NR	NR	NR	-
	Karnell MP, Melton SD, Childes JM, Coleman TC, Dailey SA and Hoffman HT [52]	NR	NR	NR	NR	NR	?
	Kazi R, De Cordova J, Singh A, Venkitaraman R, Nutting CM, Clarke P, Rhys-Evans P and Harrington KJ [102]	NR	NR	NR	NR	NR	?
	Kupfer RA, Hogikyan EM and Hogikyan ND [117]	NR	NR	NR	NR	NR	?
	Murry T, Medrado R, Hogikyan ND and Aviv JE [118]	NR	NR	NR	NR	NR	?
	Portone CR, Hapner ER, McGregor L, Otto K and Johns MM, 3rd [103]	NR	NR	NR	NR	NR	+
	Romak JJ, Orbelo DM, Maragos NE and Ekbom DC [113]	NR	NR	NR	NR	NR	+
	Tanner K, Pierce JL, Merrill RM, Miller KL, Kendall KA and Roy N [119]	NR	NR	NR	NR	NR	?
Voice Symptom Scale (VoiSS)	Deary IJ, Wilson JA, Carding PN and MacKenzie K [34]	NR	NR	NR	?	NR	NR
	Jones SM, Carding PN and Drinnan MJ [70]	NR	NR	NR	NR	NR	?
	Montgomery J, Hendry J, Wilson JA, Deary IJ and MacKenzie K [120]	NR	NR	NR	NR	NR	?
	Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen IN and Wilson JA [98]	?	-	NR	NR	NR	?
	Wilson JA, Webb A, Carding PN, Steen IN, MacKenzie K and Deary IJ [35]	+	NR	NR	NR	+	NE

Notes. Quality criteria based on those reported by Terwee and colleagues [13] and Schellingerhout and colleagues [14]

Table 10. 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) [20]	Strong (Positive result)	NR	NR	Indeterminate	Strong (Positive result)	Indeterminate
Glottal Function Index (GFI) [21]	NR	Indeterminate	NR	NR	NR	Moderate (Positive result)
Singing Voice Handicap Index (SVHI) [22]	Moderate (Negative result)	Indeterminate	NR	Indeterminate	Limited (Positive result)	Limited (Positive result)
Singing Voice Handicap Index-10 (SVHI-10) [23]	Moderate (Positive result)	Indeterminate	NR	Limited (Negative result)	NR	Limited (Positive result)
Transgender Self-Evaluation Questionnaire (TSEQ) [24]	NR	NR	NR	NR	NR	Moderate (Positive result)
Transsexual Voice Questionnaire - Male to Female (TVQ^{MtF}) [25]	Indeterminate	Strong (Positive result)	NR	NR	NR	NR
Vocal Fatigue Index (VFI) [26]	Strong (Positive result)	Indeterminate	NR	Strong (Positive result)	Moderate (Negative result)	Moderate (Positive result)
Vocal Performance Questionnaire (VPQ) [27]	Limited (Positive result)	Strong (Positive result)	NR	NR	Limited (Positive result)	Limited (Positive result)
Voice Capabilities Questionnaire (VCQ) [28]	Indeterminate	NR	NR	NR	NR	NR
Voice Disability Coping Questionnaire (VDCQ) [29]	Indeterminate	NR	NR	Limited (Positive result)	Limited (Positive result)	Indeterminate
Voice Handicap Index (VHI or VHI-30) [30]	Moderate (Positive result)	Moderate (Positive result)	NR	Indeterminate	Moderate (Positive result)	Strong (Positive result)
Voice Handicap Index-10 (VHI-10) [31]	Limited (Positive result)	NR	NR	Strong (Positive result)	Limited (Positive result)	Conflicting
Voice Rating Scale (VRS) [32]	NR	Indeterminate	NR	NR	NR	Indeterminate
Voice-Related Quality of Life (V-RQOL) [33]	Indeterminate	Limited (Positive result)	NR	NR	Moderate (Positive result)	Conflicting
Voice Symptom Scale (VoiSS) [34]	Moderate (Positive result)	Strong (Negative result)	NR	Indeterminate	Strong (Positive result)	Indeterminate

Notes. based on the levels of evidence reported by Schellingerhout and colleagues [14].