Restructuring the Vocal Fatigue Index Using Mokken Scaling: Insights Into the Complex Nature of Vocal Fatigue

***Chayadevie Nanjundeswaran, †Miriam van Mersbergen, and *Kelli Morgan,** **Johnson City and †Memphis, Tennessee*

Summary: Vocal fatigue is a frequent symptom and a debilitating condition affecting individuals with voice disorders. In spite of the various attempts to define and quantify vocal fatigue, this complex trait has not been well understood. Mokken scaling was performed on the Vocal Fatigue Index (VFI) to develop a hierarchical understanding of the latent trait of vocal fatigue. Two hundred nine patients with voice disorders completed the VFI and provided the item responses necessary to complete the Mokken scaling. Results revealed a moderately strong Mokken scale and that the VFI presents a hierarchical structure to the underlying trait of vocal fatigue. Mokken scaling contributes to the ongoing investigation to the underlying construct of vocal fatigue and may provide additional information about specific complaints within the population of those with voice disorders.

Key Words: Vocal fatigue–Vocal effort–Vocal fatigue index–Mokken scaling–Physical discomfort.

INTRODUCTION

The larynx is a unique system of the speaking mechanism that is highly susceptible to the phenomenon of fatigue, which is usually referred to as vocal fatigue.[1] Vocal fatigue is a common symptom experienced by professional voice users (ie, teachers, singers, telemarketers) who may or may not present with a voice disorder as a normal consequence of their increased vocal demands or vocal load.[2] Vocal fatigue in professional voice users varies anywhere between 18% and 33% irrespective of whether they seek treatment or not for this condition. However, vocal fatigue has gained clinical importance and is a frequent symptom and a debilitating condition affecting individuals with voice disorders.[3–5] Although vocal fatigue has been implicitly understood and defined in the literature, there is a lack of a consistent definition, making it increasingly difficult to understand the mechanisms and treatment options pertaining to it.[6,7] Vocal fatigue has been defined in a multitude of ways, thereby resulting in various methods to quantify its presence or absence. Such methods have included using patient self-reports of symptoms,[8] visual analog scales,[9] direct magnitude estimation,[10] voice adaptations as a negative consequence of prolonged voice use,[11] decrease in phonatory capacity and increased phonatory effort,[5] and possible improvement with rest.[6,12] In spite of the various attempts to define and quantify vocal fatigue, this complex trait has not been well understood. Some of the reasons for this difficulty include (a) different terminologies used to define vocal fatigue[6,7]; (b) different tasks to elicit vocal fatigue[6]; (c) varied outcome measures that may not be sensitive enough to capture the phenomenon of vocal fatigue[6]; and (d) different types of vocal fatigue including laryngeal muscular fatigue[1] (ie, peripheral fatigue), laryngeal tissue fatigue (ie, peripheral fatigue),[1] and perceived exertion of vocal effort[6,13,14] (ie, central fatigue) that may not be easily measured and quantifiable. A thorough examination of the terminologies, methods, or types of fatigue is beyond the scope of this paper, and for the purposes of this paper, we will use the term vocal fatigue to include self-reported symptoms of exertion pertaining to voice use.

Vocal Fatigue Index (VFI)

With the main assumption that vocal fatigue is a set of clinically self-perceived symptoms, the VFI was developed to reliably and consistently identify individuals presenting with this condition[12] in order to better serve the voice-disordered population who experience this debilitating condition. Further, a unified index to identify individuals with vocal fatigue will assist in the appropriate identification of assessment and treatment protocols. The current version of the VFI has 19 questions that predictably revealed that vocal fatigue is not a single construct. In its current form, the VFI is a global, retrospective, self-report of fatigue over a large segment of time (in the past few weeks). Factor analysis produced three factors underlying the trait of vocal fatigue, including factor 1, tiredness and avoidance of voice use (Tiredness and Avoidance); factor 2, physical discomfort related to voice use (Physical Discomfort); and factor 3, improvement of symptoms (or lack thereof) with rest (Symptom Improvement). The index revealed high sensitivity and specificity. Specifically, during scale validation, factors 1 and 3 were predictors of the presence or absence of vocal fatigue in a population of individuals presenting with dysphonia. However, factor 2 was not a predictor of vocal fatigue in the voice-disordered population but rather an indicator with professional users. Recent reports[15,16] have identified higher scores for factor 2 symptoms, Physical Discomfort, in teachers. It is intriguing that although logistic regression in the original version of the VFI did not indicate the importance of factor 2 in predicting vocal fatigue in the voice-disordered population, it seemed to be a significant factor in individuals with high vocal load, that is, teachers.

The fact that the VFI measures underlying constructs of patient-perceived fatigue differently depending on if one has a voice disorder or one uses his or her voice extensively suggests a more

Accepted for publication September 12, 2017.

From the *Department of Audiology and Speech-Language Pathology, East Tennessee State University, Johnson City, Tennessee; and the †School of Communication Sciences and Disorders, University of Memphis, Memphis, Tennessee.

Address correspondence and reprint requests to Chayadevie Nanjundeswaran, Department of Audiology and Speech-Language Pathology, East Tennessee State University, 254 Lamb Hall, Johnson City, TN 37614. E-mail: nanjundeswar@etsu.edu

Journal of Voice, Vol. 33, No. 1, pp. 110–114

0892-1997

https://doi.org/10.1016/j.jvoice.2017.09.008

complex underlying structure to the construct of vocal fatigue. In order to better understand the relationship between the three factors of the VFI and its individual items within the underlying trait of vocal fatigue, the Mokken Scale Analysis was utilized on an existing population with dysphonia.

Mokken scaling

Mokken scaling, a modification of Guttman scaling,[17] assumes that items in the scale reflect the underlying trait assessed *and* that each item can be placed in a hierarchy, indicating that some items reflect greater aspects of the trait than other items. This hierarchy is often referred to as "difficulty." The hierarchy of items provides insight to how an individual presents on the underlying trait of a scale. For example, an individual who presents with strong trait attributes will positively answer a difficult item, keyed "yes," in addition to also answering "yes" on items that reflect less of the underlying trait. Conversely, an individual who presents with less trait attributes will only respond to answers with low "difficulty" or those that are lower on the hierarchy.

Guttman scaling has a stricter assumption of how each item in a scale is perfectly aligned with an underlying trait.[18] The presence of the trait unquestionably predicts a "yes" key, and the absence of the factor predicts a "no" key. In contrast, Mokken scaling is more probabilistic in nature, and definite assumptions are replaced with probabilities, making it appropriate for investigations into uncharted territory, where scale development is used to investigate an underlying trait, such as the construct of vocal fatigue.[19] This probabilistic approach is common in Item Response Theory and is observed in Rasch modeling as well. However, Rasch modeling measures dichotomous items, either keyed "yes" or "no." Mokken scaling allows for items to have more than one ordered category and therefore is more appropriate for items that have more than two response choices, as seen in the VFI.[18]

In addition, Rasch models can often have criteria that are too rigid and cause items to be rejected in a scale when, in fact, the item captures a more complex relationship with the latent variable being tested.[17] For example, the item might be answered with a different set of probabilities if the underlying trait is weak (eg, in a normal, healthy population), moderate (eg, those with heavy vocal load), or strong (eg, in a disordered population) and allows for a greater understanding of the underlying trait and how it varies at different levels. This makes Mokken scaling appropriate given the observation that the factors of the VFI differ between those with voice disorders and those with high vocal load.

Other scale analysis methods such as reliability analysis (eg, Cronbach's alpha) or principal component analysis might also cause the researcher to remove items to either improve reliability or organize the data as two separate factors, ignoring the perspective of hierarchy or "difficulty."[19] If trimming or reorganizing items occurs with these other approaches, valuable information on severity is lost, forcing the scale to have a binary quality: the patient either has the problem or does not have the problem. In essence, the scale will lose its scaling ability, the ability to place individuals on a continuum of severity. For a more thorough explanation of Mokken scaling, see van Schuur.[18]

The current version of the VFI initially employed principal axis factoring with oblimin rotation to provide information on reliability and validity to measure the presence or absence of the underlying trait of vocal fatigue. Moving forward, this investigation will employ Mokken scaling to determine if there is a hierarchy within the self-perceived symptoms of vocal fatigue. Mokken scaling is measured using Loevinger's coefficient H, which is an indicator of how closely items' difficulty rates relate to each other.[16] It does this for the entire scale, H, and for each item H_i with the values between 0 (no probability) and 1 (absolute probability). If H and H_i scores are above 0.3, the scale is considered weak but valid, scores above 0.4 confirm a moderately strong scale, and scores above 0.5 indicate a strong scale. If the scale is moderate or strong, then less difficult items will provide insight into less severe symptoms of vocal fatigue and more severe symptoms of vocal fatigue would reflect more difficult items.

Mokken scaling is not new in the assessment of voice impairment indexes. Deary et al employed Mokken scaling on the Voice Symptom Scale questionnaire[20] to investigate if dysphonia symptoms were hierarchical. They found that symptoms of dysphonia progressed from physical symptoms to emotional symptoms. In doing so, the authors not only continued the reliability testing of the scale but also discovered valuable information about the latent trait the scale purports to measure. Mokken scaling has this dual benefit and is appropriate for the VFI.

Purpose of the study

The general aim of the present study is to establish the relationship among items in the current VFI and develop a hierarchical understanding of the latent trait of vocal fatigue. Additionally, this study intends to understand where factor 2, Physical Discomfort, falls within the hierarchy of the latent trait of vocal fatigue. Finally, this study will provide potential insight to the development of additional perceived exertion scales directed as more immediate ratings, such as immediately after a vocal task (as observed in classic stress test compared with retrospective perceptions "in the last few weeks").

METHODS

Participants and design

This retrospective study employed data obtained from a combined sample of 233 individuals presenting with dysphonia from two different sites to run Mokken analysis on a larger data set. The first set of data was utilized from the original validation study of the current version of the VFI completed by individuals with dysphonia at the University of Pittsburgh Voice Center and Vanderbilt Voice Center (n = 80).[12] The second set of data was obtained from the clinical setting at Johnson City Ear, Nose and Throat Associates (n = 153). Institutional Review Board approval was obtained prior to data analysis. The participants presented with multiple dysphonia diagnoses, and to capture the largest, most representative sample, no attempts were made to exclude any diagnoses. Of the 233 individuals, 26 were removed from the analysis due to missing data. Therefore, Mokken scale analysis was completed on a total of 207 participants including 132 females and 75 males in total.

VFI

The VFI is a 19-question self-reported questionnaire that targets symptoms related to vocal fatigue. Each item has a stem question with a five-choice Likert-type response scale: never, almost never, sometimes, almost always, and always. The VFI has three factors, each identifying a different aspect of vocal fatigue including Tiredness and Avoidance, Physical Discomfort, and Symptom Improvement (with rest). As stated in the Introduction, the index is highly reliable and has a high sensitivity and specificity for identifying the presence or absence vocal fatigue. However, it is unknown if the items form a natural, consistent order of severity within the underlying trait of vocal fatigue.

Mokken scaling

Data were entered into a spreadsheet (*Microsoft Excel 2016*; Microsoft Corp., Seattle, Washington), and as indicated, 26 participants with missing data were removed from further analysis. The data were then converted and processed for Mokken scaling using the *RStudio software* (Version 3.3.2) on the PC version (R Foundation for Statistical Computing, Vienna, Austria). Using Mokken scaling criteria, data were divided into two separate clusters by the R program using the aisp (FileR) command. The number of scales, their reliability (Rho), and the invariant item ordering (H_i) were recorded. Further analysis of monotone homogeneity and double monotonicity were carried out to identify any outliers or violations of invariant item ordering. The data were then listed in order of difficulty in its cluster based on the mean score.

RESULTS

From the current version of 19 questions on the VFI, two clusters were appropriate for Mokken scaling and comprised two separate Mokken scales. Cluster 1 represents a group of 16 questions that correlate and form a hierarchy of difficulty, while cluster 2 included only three correlating questions and did not reveal a strong Mokken scale. After checking for monotone homogeneity and double monotonicity, a reliable (Rho = 0.952) and moderately strong ($H^t = 0.464$) scale was found in cluster 1. Table 1 shows a scale for cluster 1 that increases in difficulty from "4. My voice gets hoarse with voice use" (mean = 3.014) to "12. I experience pain in the neck at the end of the day with voice use" (mean = 1.256).

Moving from items with the least difficulty (highest mean) to those with the most difficulty (lowest mean), the scale traces a process of first reporting symptoms of tiredness (eg, "4. My voice gets hoarse with voice use"; "2. My voice feels tired when I talk more") to avoidance (eg, "1. I don't feel like talking after a period of voice use"; "6. I tend to generally limit my talking after a period of voice use") and then progresses to physical discomfort (eg, "15. My throat aches with voice use"; "12. I experience pain in the neck at the end of the day with voice use").

Although cluster 2 did not meet the criteria for a valid Mokken scale ($H^t = 0.019$), it is important to mention that this cluster includes all of the questions of factor 3 of the VFI, Symptom Improvement (with rest), with a high correlation between items (Rho = 0.844). It was expected that these items would score separately from the remaining questions. Table 2 shows the hierarchy of question difficulty within cluster 2. Table 3 is presented with the items on the VFI in the hierarchical order as obtained using Mokken scaling. However, this table should not replace the current version of the VFI for research or clinical purposes because a change in order may affect the psychometric properties, specifically the item bias.[21]

TABLE 1.
Cluster 1 Indicating the Reorganization of Factors 1 (Tiredness and Avoidance) and 2 (Physical Discomfort) of the VFI ($H^t = 0.464$; Rho = 0.952)

Question	Mean	H_i
4. My voice gets hoarse with voice use.	3.014	0.57
2. My voice feels tired when I talk more.	2.816	0.66
11. My voice feels weak after a period of voice use.	2.705	0.66
3. I experience increased sense of effort with talking.	2.676	0.63
10. I find it difficult to project my voice with voice use.	2.632	0.55
5. It feels like work to use my voice.	2.512	0.65
9. It is effortful to produce my voice after a period of voice use.	2.473	0.64
1. I don't feel like talking after a period of voice use.	2.411	0.66
6. I tend to generally limit my talking after a period of voice use.	2.406	0.61
7. I avoid social situations when I know I have to talk more.	1.913	0.57
14. My voice feels sore when I talk more.	1.734	0.50
15. My throat aches with voice use.	1.580	0.61
13. I experience throat pain at the end of the day with voice use.	1.498	0.59
16. I experience discomfort in my neck with voice use.	1.391	0.60
8. I feel I cannot talk to my family after a workday.	1.280	0.53
12. I experience pain in the neck at the end of the day with voice use.	1.256	0.60

Notes: Items at the top are endorsed first ("easiest items"), and those at the bottom are endorsed only when the top items are endorsed ("difficult items").

TABLE 2.
Cluster 2 Representing Factor 3 (Symptom Improvement) of the VFI ($H^t = 0.019$; Rho = 0.844)

Question	Mean	H_i
17. My voice feels better after I have rested.	2.304	0.65
18. The effort to produce my voice decreases after rest.	2.135	0.65
19. The hoarseness of my voice gets better with rest.	2.198	0.63

Notes: Items in this cluster did not qualify for further Mokken criteria.

TABLE 3.
Hierarchical Ordering of VFI Items Based on Mokken Scaling

VFI-Mokken Scaling Hierarchy					
Cluster 1:					
1. My voice gets hoarse with voice use.	0	1	2	3	4
2. My voice feels tired when I talk more.	0	1	2	3	4
3. My voice feels weak after a period of voice use.	0	1	2	3	4
4. I experience increased sense of effort with talking.	0	1	2	3	4
5. I find it difficult to project my voice with voice use.	0	1	2	3	4
6. It feels like work to use my voice.	0	1	2	3	4
7. It is effortful to produce my voice after a period of voice use.	0	1	2	3	4
8. I don't feel like talking after a period of voice use.	0	1	2	3	4
9. I tend to generally limit my talking after a period of voice use.	0	1	2	3	4
10. I avoid social situations when I know I have to talk more.	0	1	2	3	4
11. My voice feels sore when I talk more.	0	1	2yg	3	4
12. My throat aches with voice use.	0	1	2	3	4
13. I experience throat pain at the end of the day with voice use.	0	1	2	3	4
14. I experience discomfort in my neck with voice use.	0	1	2	3	4
15. I feel I cannot talk to my family after a workday.	0	1	2	3	4
16. I experience pain in the neck at the end of the day with voice use.	0	1	2	3	4
Cluster 2:					
17. My voice feels better after I have rested.	0	1	2	3	4
18. The effort to produce my voice decreases with rest.	0	1	2	3	4
19. The hoarseness of my voice gets better with rest.	0	1	2	3	4

Notes: Please note that this version of the VFI is not to replace the original version of the VFI for clinical or research purposes.

DISCUSSION

Mokken scaling was employed to understand the relationship between VFI items and their hierarchy within this complex trait of vocal fatigue. Of the 19 items appropriate for Mokken scaling, two clusters (cluster 1 and cluster 2) with separate Mokken scales were formed. According to the criteria for judging Mokken scale, cluster 1 of the VFI indicates a moderately strong scale and strong Rho values indicate the reliability of the scale.

Within cluster 1 obtained using Mokken scaling, items on factors 1 (Tiredness and Avoidance) and 2 (Physical Discomfort) of the original VFI validation are combined and reorganized. Within this reorganization, it is evident that at the mild end of this hierarchy, participants report Tiredness, progressing to Avoidance, and finally at the severe end, Physical Discomfort symptoms. This is compared with the original validation that combined Tiredness and Avoidance together and treated Physical Discomfort as a separate but equal factor. This reorganization is robust with the exception of one item ("I feel I cannot talk to my family after a workday."). Mokken scaling captures the unique contribution of this item within this hierarchy. Although it falls under the avoidance factor, this item is pulled out and made more difficult. This could be attributed to the fact that this item provides high motivation to talk to one's family. This way Mokken scaling assists in reorganizing the motivational structure of communication within the underlying trait of fatigue. Because there is only one item on this motivational factor, Mokken scaling uncovers the need for possible additional items addressing this influence. Cluster 2 comprised the three Symptom Improvement questions. Although cluster 2 did not meet adequate criteria for Mokken scaling, the reliability within items was high. This cluster reinforces the notion that voice rest and recovery, in relation to vocal fatigue, is an important influence within the construct of vocal fatigue[6] and requires more attention.

Mokken scaling further established the reliability and validity of the existing VFI. Additionally, the reorganization of the factors in the hierarchy provides insight to the complex nature of the underlying trait of vocal fatigue. The current reorganization is based on the population studied individuals with dysphonia with a broad diagnosis of voice disorders. However, questions remain about this reorganization. Are Fatigue and Physical Symptoms orthogonal, sequential, or embedded? In this population, it appears that Fatigue and Physical Symptoms are sequential and that one indicates a lesser problem than the other. However, would other populations show a different pattern? As mentioned earlier, those without voice disorders might show a different scaling hierarchy from teachers. If there is a different "teacher hierarchy," where vocal fatigue is a highly prevalent experience,[22] then comparisons between these hierarchies might be useful. This was observed in the aforementioned reports where teachers reported Physical Discomfort as a more likely symptom than those with voice disorders. The differences in how a given population responds to the VFI might provide additional diagnostic insight. For example, a teacher who responds to VFI questions more similarly to those with voice disorders than other teachers might be considered an at-risk individual. Furthermore, in some populations (ie, singers, varied diagnoses of voice-disordered population), the items themselves might be clustered differently, allowing the clinician and the researcher to understand the unique experience of vocal fatigue in that population. These data will provide insight to directing the choice of appropriate preventative or treatment approaches to alleviating the symptom of vocal fatigue in different populations.

Another useful outcome of Mokken scaling is the guidance it provides in developing appropriate anchors for *in vivo* vocal fatigue measures such as the Borg CR10. In this measure, an individual engages in a vocal activity and immediately responds to his or her perceived level of fatigue for the particular task. A challenging part of this type of effort scaling is understanding the appropriate anchors for minimal and maximal measures of vocal fatigue.[13,22] Understanding overall hierarchies

in measures such as the VFI might clarify the types of anchors to use for each level of the Borg CR10 scale. Additionally, using the hierarchies of the VFI gathered from large data sets might define ubiquitously experienced phenomena that would allow anchors to be more universal, allowing scores to be compared within and among individuals.[13,22] Finally, knowing the more difficult items for individuals on the VFI might guide the task choice to elicit *in vivo* effort ratings. For example, would prolonged voice use be a more appropriate fatigue-inducing activity than projected voice use in a certain population? According to this research, perceived discomfort was rated more "difficult," so a task that might elicit discomfort (as might be observed in loud talking) would uncover vocal fatigue more quickly.

Limitations of this study

This study presents several limitations. First, the study was conducted on a heterogeneous population of individuals with voice disorders. In order to employ Mokken scaling as a differential diagnostic tool across voice disorders, this population should be theoretically categorized and analyzed separately (eg, those with benign vocal fold lesions compared with those with unilateral vocal fold paralysis). Despite this limitation, the study shows a robust hierarchy indicating a common experience within the trait of vocal fatigue among this heterogeneous population. Second, along these same lines, this study used 209 participants for analysis. Larger numbers with more distinct categories will be necessary for future investigations. Lastly, it is important to mention that the VFI is a 19-item scale. Although this scale offers excellent specificity and sensitivity in determining whether someone experiences vocal fatigue, more scale items might be necessary to investigate signature response patterns among different populations.[14] Despite this, Mokken scaling with the current items revealed a moderately strong hierarchy.

Future directions

Future directions for the use of Mokken scaling are exciting. Results from this approach of scaling provides (a) further support to the reliability and validity of the established VFI, (b) a concept of hierarchy/progression of symptoms from items of low to high difficulty within the trait of vocal fatigue, and finally (c) the complex nature of this underlying trait of vocal fatigue. However, it has to be acknowledged that the current hierarchy exists for the general population of individuals with dysphonia with a multitude of voice diagnoses. Further work is required to see if Mokken scaling hierarchy will vary based on the population of choice (ie, in professional voice users). Additional Mokken scaling on other populations, including professional voice users or vocally healthy individuals, could provide further insight to the development of an analysis or scoring protocol based on a Mokken hierarchy that would allow for further understanding of the complex, heterogeneous concept of vocal fatigue.

Conclusions

Mokken scaling provided additional insight to the information provided by the VFI. From the original VFI validation, Mokken scaling combined factors 1 and 2 and reordered its structure. Initial scaling provides additional insight to the underlying factor of vocal fatigue and may provide guidance to future studies investigating how vocal fatigue differs among various populations.

Acknowledgments

The authors wish to thank the Pittsburgh Voice Center, Vanderbilt Voice Center, and Johnson City Ear, Nose and Throat Associates for their cooperation in obtaining VFIs on their patients. The second author would like to thank Kimberly Steinhauer on inspiring her to explore vocal effort and fatigue via multiple psychometric methods.

REFERENCES

1. Sander EK, Ripich DE. Vocal fatigue. *Ann Otol Rhinol Laryngol*. 1983;92:141–145.
2. Titze IR. Toward occupational safety criteria for vocalization. *Logoped Phoniatr Vocol*. 1999;24:49–54.
3. Colton RH, Casper JK, Leonard R. *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. 3rd edition. Baltimore, MD: Lippincott Williams & WIlkins; 2006.
4. Gotaas C, Starr CD. Vocal fatigue among teachers. *Folia Phoniatr (Basel)*. 1993;45:120–129.
5. McCabe DJ, Titze IR. Chant therapy for treating vocal fatigue among public school teachers: a preliminary study. *Am J Speech-Lang Pat*. 2002;11:356–369.
6. Solomon NP. Vocal fatigue and its relation to vocal hyperfunction. *Int J Speech-Lang Pa*. 2007;10:1–13.
7. Welham NV, Maclagan MA. Vocal fatigue: current knowledge and future directions. *J Voice*. 2003;17:21–30.
8. Kitch JA, Oates J. The perceptual features of vocal fatigue as self-reported by a group of actors and singers. *J Voice*. 1994;8:207–214.
9. Solomon NP, Glaze LE, Arnold RR, et al. Effects of a vocally fatiguing task and systemic hydration on men's voices. *J Voice*. 2003;17:31–46.
10. Chang A, Karnell MP. Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: a study of vocal fatigue. *J Voice*. 2004;18:454–466.
11. Vilkman E. Occupational safety and health aspects of voice and speech professions. *Folia Phoniatr Logo*. 2004;56:220–253.
12. Nanjundeswaran C, Jacobson BH, Gartner-Schmidt J, et al. Vocal Fatigue Index (VFI): development and validation. *J Voice*. 2015;29:433–440.
13. Baldner EF, Doll E, van Mersbergen MR. A review of measures of vocal effort with a preliminary study on the establishment of a vocal effort measure. *J Voice*. 2012;29:530–541.
14. Furr M. Core principles, best practices and an overview of scale construction. In: *Scale construction and psychometrics for social and personality psychology*. 2011:4–15.
15. Dietrich M, Glandon E, Williams K, et al. The role of a weak vocal constitution in risk for voice problems in student teacher. Poster Session, Board 531. ASHA Convention: Denver, CO. 2015.
16. Banks R, Hunter EJ, Graetzer S, et al. Vocal risks and vocal fatigue in school teachers. 2015 American Speech-Language-Hearing Association (ASHA) Convention, Denver, CO. 2015.
17. Sijtsma K, Molenaar WI. *Introduction to Nonparametric Item Response Theory*. Thousand Oaks, CA: Sage; 1987.
18. van Schuur WH. Mokken scale analysis: between the Guttman scale and parametric item response therapy. *Polit Anal*. 2003;11:139–163.
19. Stochl J, Jones PB, Croudace TF. Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers. *BMC Med Res Meth*. 2012;12:74.
20. Deary IJ, Wilson JA, Carding PN, et al. VoiSS: a patient derived voice symptom scale. *J Psychosom Res*. 2003;54:483–489.
21. Crocker L, Algina J. *Introduction to Classical and Modern Test Theory*. Fort Worth, TX: Harcourt; 1986:378–379.
22. Smith E, Kirchner HL, Taylor M, et al. Voice problems among teachers: differences by gender and teaching characteristics. *J Voice*. 1998;12:328–334.