Screening for Voice Disorders in Older Adults (*Rastreamento de Alterações Vocais em Idosos—* RAVI)—Part II: Validity Evidence and Reliability

*,†Leandro de Araújo Pernambuco, ‡,§,||Albert Espelt, †Erika Beatriz de Morais Costa, and *Kenio Costa de Lima,
*†§*Rio Grande do Norte, Brazil, and ‡||§Barcelona, Bellaterra, Spain*

Summary: Purpose. To determine the validity evidence based on the internal structure and relations to other variables and the reliability of an epidemiologic questionnaire for screening older adults with voice disorders.

Study Design. This is a prospective, nonrandomized, cross-sectional, validation study.

Methods. To assess the validity evidence based on the internal structure, 160 older adults of both sexes, either community-dwelling or institutionalized, completed the "Rastreamento de Alterações Vocais em Idosos" (RAVI; "Screening for Voice Disorders in Older Adults"). The data were analyzed using item-total and interitem correlations, principal component analysis, confirmatory factor analysis, and differential item functioning. For validity evidence based on relations to other variables, measures of concurrent and discriminant validity were calculated according to the self-reported number of factors associated with voice disorders. Reliability was assessed using a test-retest procedure conducted with 121 older adults with the same characteristics as the previous sample. In this stage, the reproducibility, internal consistency, and measurement error of the instrument were analyzed.

Results. The RAVI showed adequate validity evidence based on the internal structure and relations to other variables, as well as good reliability. This indicates that the results produced by the instrument are valid and reliable.

Conclusions. The RAVI is a self-reported outcome questionnaire that yields valid and reliable responses for epidemiologic detection of voice disorders in older adults and can be easily administered by any health care provider. Further ongoing studies will assess accuracy measures and cutoff values of the RAVI.

Key Words: Voice–Voice disorders–Dysphonia–Aged–Aging–Health of the elderly–Epidemiology–Validation studies.

INTRODUCTION

Older adults' greater exposure to voice disorders and the fast-paced growth of this stratum of the population worldwide indicate an urgent need for instruments that can substantiate the demographic information related to this health condition.[1] Such is the purpose of the "Rastreamento de Alterações Vocais em Idosos" (RAVI; "Screening for Voice Disorders in Older Adults"), a simple, inexpensive, short, and easy-to-administer questionnaire designed for population health surveys intended to map the prevalence or incidence of voice disorders in older adults.

Novel instruments such as the RAVI should ensure that interpretations of the results accurately reflect the intended construct and should be reliable in terms of question consistency, sustained reproducibility, and control of measurement errors.[2,3] This can be done by following the series of steps to secure psychometric properties laid out in the "Standards for Educational and Psychological Testing,"[3,4] a document that

provides criteria for the development and evaluation of tests and testing practices, and guidelines for assessing how valid interpretations of test scores are for the test's intended use.[5]

The "Standards for Educational and Psychological Testing"[3] provides five sources of validity evidence, those based on test content, response processes, internal structure, relation to other variables, and test consequences. Validity evidence based on test content and response processes, both of which are central to the syntactic, semantic, and contextual adequacy of the questions in relation to the construct, was presented in part I of this study.[6] Part II presents the validity evidence based on the internal structure and relations to other variables and the results of a reliability analysis.

Validity evidence based on the internal structure covers the following: (1) the relationship between the instrument questions (ie, to what extent one question of the RAVI is correlated to all the other questions); (2) the relationship of the questions with the total scores (ie, how strong the questions correlate to the final score); (3) to what extent the questions are related to the aim of the questionnaire; and (4) whether different groups of interviewers (eg, according to sex, race, education level) with similar abilities have, on average, systematically different responses to a particular question.[3,5,7–9]

Validity evidence based on relation to other variables refers to whether the instrument scores have consistent linkages to external variables with a similar or dissimilar construct[3,5,7] (eg, whether a voice-related quality of life questionnaire score relates to a depressive symptom measure). This type of evidence addresses the degree to which the relationships between the instrument scores and external variables are consistent with the

construct underlying the proposed instrument. At this stage, it is possible to analyze the relationship between the scores of the new test and a relevant criterion defined by the researchers.[3] This analysis can be performed concurrently (ie, test and criterion evaluated at the same time) or predictively (ie, how much a test can predict information about a criterion that will be obtained later).[2,3] Furthermore, it is possible to investigate evidence of convergent validity (ie, comparing the new instrument with a previous instrument attempting to measure the same construct) or discriminant validity (ie, comparing the new instrument with an instrument that measures a different construct or comparing the results of the new instrument between two groups of people with specific characteristics).[5]

Reliability refers to the internal consistency, reproducibility, and control of measurement errors.[2,3] Internal consistency refers to what extent the scores are truly dependent on the questions of the instrument (homogeneity); measurement error refers to the influence of random and systematic errors inherent to the respondent (eg, forget to answer one question) or instrument variability (eg, extent of the questions); and reproducibility refers to the consistency of the responses to the questions when these are presented under the same conditions, to the same individuals, at separate times.[2,3,5]

To date, there are no instruments designed to identify voice disorders in the older adult population, especially from an epidemiologic perspective. The instruments that have been frequently used in voice research are intended to assess other dimensions, such as voice handicap,[10] therapeutic effectiveness,[11] and voice-related impact on quality of life.[12] These instruments were not originally designed considering the specificities of older adults' voices[6]; in addition, there are problems in their psychometric properties.[13] Thus, the objective of the present study was to gather validity evidence based on the internal structure and relation to other variables as well as to verify the reliability of the RAVI.

METHODS

Throughout its stages, the present study included Brazilian older adults (aged ≥60 years according to the World Health Organization criteria for developing countries), both sexes, randomly selected from communities and long-term care institutions. We administered the RAVI to a representative sample of the target population (any older person with or without voice complaints). We decided not to consider vocal quality status of the older adult as a variable because psychometric properties evaluated at this stage are intrinsic to the test and not to the person. Furthermore, RAVI is an epidemiologic questionnaire, which means that any older person of the community is a target, regardless of its vocal quality status. Respondents' vocal quality status is being considered in ongoing studies on the clinical consistency and cutoff points of the RAVI.

Respondents were excluded if they were unable to complete the questionnaires; could not understand or execute simple instructions; had a psychiatric, neurologic, neuromuscular, or neurodegenerative disorder; had impaired consciousness; were unable to remain in a sitting position; had a total or partial laryngectomy; or had a tracheostomy. Respondents were checked for these exclusion criteria via respondents' self-

reported conditions, caregiver reports, notes in medical records, or the evaluator's perceptions.

Validity evidence based on internal structure

At this stage, 160 older adults aged 60–98 years (mean, 72.04 ± 6.38 years), who were predominantly female (n = 111; 69.4%), were randomly selected for participation. The sample size conformed to the traditionally recommended ratio of a minimum of 10 participants per item or question[14] for conducting a principal component analysis (PCA). The purpose of the PCA was to check if the number of questions could be reduced by examining the question interrelationships. To ascertain whether a PCA could be performed, the correlation matrix was visually examined for a favorable number of correlations equal to or exceeding 0.30, a Kaiser-Meyer-Olkin (KMO) of 0.60 or higher for measure sampling adequacy of the whole test, a measure sampling adequacy of individual questions of 0.5 or higher, a significant Bartlett's test of sphericity (at 5% significance level), and partial correlations of 0.7 or less.[15]

The components were extracted using an orthogonal varimax rotation. The communalities analysis indicated how much of the variance of each variable was explained by the model estimated by the PCA; values of ≥0.50 were considered acceptable.[15] Components were selected using Kaiser normalization, which considers eigenvalues exceeding 1.0 for inclusion. Subsequently, the rotated component matrix was analyzed, with only component loadings >0.5 being retained.[15] The components were then defined along with the percentage of variance explained by each component. A confirmatory factor analysis (CFA) was conducted to validate the PCA model. In addition, the item-total and interitem correlations were analyzed at this stage; values of ≥0.3 were deemed acceptable.[2] Differential item functioning (DIF)[8] was detected using simple bivariate logistic regression on the basis of the variables sex and education level.

Validity evidence based on relations to other variables

Evidence of criterion-related concurrent validity was ascertained using the same sample of 160 individuals investigated in the previous stage. The established criterion was the correlation between the final RAVI score and the number of factors associated with voice disorders found in the literature[16] (Figure 1). It was hypothesized that older adults with more factors associated with voice disorders would have higher scores (ie, poorer outcomes) on the RAVI. We analyzed the factors associated with voice disorders mentioned in a widely cited North American epidemiologic study about aging voices[16] and selected the factors that were statistically significant in the original study or that we considered theoretically relevant by consensus. No studies with similar characteristics to the North American study were found in Brazil.

The correlation between the RAVI scores and the number of associated factors reported by the participants was tested using the Pearson correlation coefficient. We split participants into two groups of high and low number of associated factors via a median split. Student's *t* test was used to assess the difference in the means of the total RAVI scores between the two groups