This is a draft of a paper that has been accepted for publication in Ear and Hearing. Please see, and cite if relevant, the final published version. doi: 10.1097/AUD.0000000000001369 (note the doi is not yet active, 3/21/2023. Check back in 6-8 weeks).

Development and Validation of a Brief Version of the Vanderbilt Fatigue Scale for Adults: The VFS-A-10

Benjamin W.Y. Hornsby[1], Stephen Camarata[1,2], Sun-Joo Cho[2], Hilary Davis[1], Ronan McGarrigle[3] and Fred H. Bess[1]

Affiliation:

[1]Vanderbilt University School of Medicine, Department of Hearing & Speech Sciences, Nashville, TN, United States

[2]Vanderbilt University, Peabody College, Department of Psychology & Human Development, Nashville, TN, United States

[3]Department of Psychology, University of Bradford, Bradford, UK

Financial disclosures/conflicts of interest: This work was supported by a grant from Starkey Inc (Hornsby, PI), by the NIH National Institute on Deafness and Other Communication Disorders (NIDCD) Grant #R21DC012865 (Hornsby, PI), by an NICHD Grant P30HD15052 to the Vanderbilt Kennedy Center for Research on Human Development, and by a Vanderbilt Institute for Clinical and Translational Research grant (UL1 TR000445 from NCATS/NIH). The opinions expressed are those of the authors and do not represent the views of the NIH or other institutions. The funding organization had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the decision to submit the article for publication; or in the preparation, review, or approval of the article.

An electronic version of the 10-item Vanderbilt Fatigue Scale for Adults (VFS-A-10) described in this paper is available for download at https://www.vumc.org/vfs

All correspondence should be addressed to:
Benjamin Hornsby, Department of Hearing & Speech Sciences, Vanderbilt University School of Medicine, Vanderbilt Bill Wilkerson Center, 1215 21st Avenue South, MCE South Tower, Room 8310 Nashville, Tennessee 37232-8718, United States. E-Mail: ben.hornsby@vumc.org

Abstract

Objectives: Listening-related fatigue can be a significant problem for adults who struggle to hear and understand, particularly adults with hearing loss. However, valid, sensitive, and clinically useful measures for listening-related fatigue do not currently exist. The purpose of this study was to develop and validate a brief clinical tool for measuring listening-related fatigue in adults.

Design: The clinical scale was derived from the 40-item version of the Vanderbilt Fatigue Scale for Adults (VFS-A-40), an existing, reliable, and valid research tool for measuring listening-related fatigue. The study consisted of two phases. Phase 1 (N = 580) and Phase 2 (N = 607) participants consisted of convenience samples of adults recruited via online advertisements, clinical records review, and a pool of prior research participants. In Phase 1, results from item response theory (IRT) analyses of VFS-A-40 items were used to identify high quality items for the brief (10-item) clinical scale: the VFS-A-10. In Phase 2, the characteristics and quality of the VFS-A-10 were evaluated in a separate sample of respondents. Dimensionality was evaluated using exploratory factor analyses (EFA) and item quality and characteristics were evaluated using IRT. VFS-A-10 reliability and validity were assessed multiple ways. IRT reliability analysis was used to examine VFS-A-10 measurement fidelity. In addition, test-retest reliability was assessed in a subset of Phase 2 participants (n = 145) who completed the VFS-A-10 a second time approximately one month after their initial measure (range 5-90 days). IRT differential item functioning (DIF) was used to assess item bias across different age, gender, and hearing loss subgroups. Convergent construct validity was evaluated by comparing VFS-A-10 responses to two other generic fatigue scales and a measure of hearing disability. Known-groups validity was assessed by comparing VFS-A-10 scores between adults with and without self-reported hearing loss.

Results: EFA suggested a unidimensional structure for the VFS-A-10. IRT analyses confirmed all test items were high quality. IRT reliability analysis revealed good measurement fidelity over a wide range of fatigue severities. Test-retest reliability was excellent (r_s = .88, collapsed across participants). IRT DIF analyses confirmed the VFS-A-10 provided a valid measure of listening-related fatigue regardless of respondent age, gender, or hearing status. An examination of associations between VFS-A-10 scores and generic fatigue/vigor measures revealed only weak-to-moderate correlations (Spearman's correlation coefficient r_s = -.36 to .57). Stronger associations were seen between VFS-A-10 scores and a measure of perceived hearing difficulties (r_s = .79 to .81) providing evidence of convergent construct validity. In addition, the VFS-A-10 was more sensitive to fatigue associated with self-reported hearing difficulties than generic measures. It was also more sensitive than generic measures to variations in fatigue as a function of degree of hearing impairment.

Conclusions: These findings suggest that the VFS-A-10 is a reliable, valid, and sensitive tool for measuring listening-related fatigue in adults. Its brevity, high sensitivity, and good reliability make it appropriate for clinical use. The scale will be useful for identifying those most affected by listening-related fatigue and for assessing benefits of interventions designed to reduce its negative effects.

INTRODUCTION

The psychosocial consequences of hearing loss are significant and diverse. One important, but largely understudied, issue associated with hearing loss is listening-related fatigue. Listening-related fatigue, like the broader construct of subjective fatigue, can be defined multiple ways (see Hornsby, Naylor, & Bess, 2016, for review). Although there is no gold standard definition, fatigue is often defined and measured subjectively as a mood state—a general feeling of tiredness or weariness whose magnitude can range from mild to severe. Fatigue magnitude can be influenced by many factors; however, fatigue is often thought to be a direct consequence of the sustained application of high levels of effort towards a task (Hockey, 2013; Hornsby, Naylor, & Bess, 2016). *Listening*-related fatigue is a type of fatigue associated with the sustained application of effort towards a *listening task* (Bess & Hornsby, 2014; Pichora-Fuller et al., 2016; Davis et al., 2021a; Hornsby et al., 2021). For people with hearing loss, listening, particularly in challenging acoustic environments, can require substantial mental effort and thus increase the likelihood of developing listening-related fatigue (Hornsby, 2013; Alhanbali et al., 2017; Holman et al., 2019; Davis et al., 2021a; Davis et al., 2021b).

When listening-related fatigue is recurrent and/or severe it can have significant negative effects on quality of life in people with hearing loss (Holman et al., 2019; Davis et al., 2021a; Davis et al., 2021b). Davis and colleagues (2021a,b) conducted focus groups and interviews with adults and children with hearing loss to better understand their perspective of listening-related fatigue and its impact on their daily lives. While some specifics varied between adults and children based on their environmental settings (e.g., work versus school), substantial similarities were also observed. Study participants reported that listening-related fatigue could have physical, cognitive, social, and emotional manifestations that impacted their quality of life. For example, general feelings of tiredness resulting from listening difficulties could lead to challenges maintaining active engagement in listening situations (e.g., in social, school, or work situations). Adults and children reported using a variety of positive and negative coping strategies, such as withdrawing from social settings or "shutting down" (i.e., stop actively engaging in the listening), to recover from or avoid these fatigue effects. Consistent with these reports, data from the broader fatigue literature suggests that fatigued adults are less productive at work, more prone to accidents, more socially isolated, and more likely to suffer from depression than their non-fatigued peers (Amato et al., 2001; Eddy & Cruz, 2007; Ricci et al., 2007). Likewise, children experiencing severe, recurrent fatigue are more likely to miss school or do poorly in school compared to their non-fatigued peers (Hockenberry-Eaton et al., 1999; Gaba & Howard, 2002; Ravid et al., 2009).

Given the significant negative effects of fatigue on quality of life, a variety of subjective measures have been developed to quantify the frequency and severity of its effects (see Dittner, 2004 & Whitehead, 2009, for reviews). Broadly speaking, the fatigue scales described in the literature can be categorized as generic or disease-specific measures. Generic instruments can be used to measure fatigue associated with a wide range of chronic health conditions. For example, the Pediatric Quality of Life Multidimensional Fatigue Scale (PedsQL-MFS; Varni et al., 2002) has been used to assess childhood fatigue associated with health conditions such as cancer, multiple sclerosis, diabetes, irritable bowel syndrome, rheumatoid arthritis, obesity, short stature, and hearing loss (see Hornsby et al., 2017, for discussion). Other scales are designed to assess fatigue associated with specific disease conditions (e.g., cancer) or populations (e.g., athletes). For example, the Revised Piper Fatigue Scale (Piper et al., 1998) was designed specifically to assess cancer-related fatigue, although it can be used to examine fatigue associated with other health conditions, such as post-polio fatigue (Strohschein et al., 2003). Both generic and disease-specific measures are useful; however, well-designed, disease-specific instruments may be more sensitive to variations in a specific test construct (e.g., cancer-related fatigue) than generic measures, at least when used with their target population (e.g., Dwyer et al., 2019; Holman, Drummond, & Naylor, 2021; Patrick & Deyo, 1989). For example, Dwyer et al. (2019) compared fatigue ratings in a small group of younger adult cochlear implant and hearing aid users and an age-matched control group without hearing loss. Fatigue ratings were obtained using a generic fatigue measure, the Profile of Mood States (POMS; McNair et al., 1971), and using three, non-validated, items to ask specifically about listening-related fatigue problems. POMS ratings were not significantly different between groups whereas ratings based on items targeting listening-related fatigue revealed large and significant differences. Work like this highlights the importance of using measures designed to target the problem of interest.

Importantly, until recently there were no validated scales specifically developed to assess listening-related fatigue. The pediatric and adult versions of the Vanderbilt Fatigue Scales (VFSs) were designed to fill this gap. A reliable and valid self-report measure is particularly important given prior work has found limited associations between self-reported fatigue and traditional audiometric measures such as pure-tone average hearing threshold (Alhanbali et al. 2017; Hornsby & Kipp, 2016). There are three pediatric versions of the Vanderbilt Fatigue Scales (the VFS-Peds)— a child self-report version (VFS-C), a parent proxy-report version (VFS-P), and a teacher proxy-report version (VFS-T). Details regarding the development and validation of these pediatric scales are provided in Hornsby et al. (2022). Relevant to this paper, a 40-item version of the VFS has been developed for the adult population—the VFS-A-40. The VFS-A-40 was designed to measure long-term listening-related fatigue in adults (i.e., fatigue experienced over a typical week). The scale has good test reliability, comparability across subgroups, and construct validity (Hornsby et al., 2021).

To enhance its construct validity, VFS-A-40 test items were developed to assess the cognitive, physical[1], social, and emotional characteristics of listening-related fatigue as described by adults with hearing loss (Davis et al., 2021a). As such, the scale allows for the calculation of four subscale scores reflecting these various domains. Subscale scores are calculated based on responses to 10 items/domain. However, exploratory factor analyses (EFA) completed during the development of the VFS-A-40 suggested that all subscale scores reflected a single underlying dimension of listening-related fatigue (Hornsby et al., 2021), thus a total listening-related fatigue score based on responses to all test items can also be calculated. Despite associations between subscale scores, Hornsby and colleagues (2021) argued that an examination of subscale scores may still be useful in certain clinical and/or research settings. For example, McGarrigle et al., (2021a) used the VFS-A-40 to investigate the effect of age on listening-related effort and fatigue in a sample of young and older adults without self-reported hearing loss. Interestingly, they found significant age-related effects, but only for the social domain of listening-related fatigue. No age-related effects were observed for total (unidimensional) scores or for scores in any other domains.

Recent research has found the VFS-A-40 to be sensitive to variations in listening-related fatigue due to hearing aid use (Holman, Drummond, & Naylor, 2021) and age (McGarrigle et al., 2021b). For example, Holman and colleagues (2021) examined the impact of hearing aid use on

[1] It is worth noting that VFS-A-40 items targeting the physical domain of listening-related fatigue do not query respondents about physical activities that lead to fatigue. Rather they ask or comment on general feelings of tiredness (e.g., I feel “tired” or “worn out”) or actions that respondents engage in that can be physically observed (e.g., “I schedule my day to avoid getting tired…” or “…I go to bed early” due to listening challenges).

fatigue in a group of first-time hearing aid users. They used two generic fatigue scales and the VFS-A-40 (a disease-specific scale) to measure fatigue prior to receiving hearing aids and then again two weeks, 3 months, and 6 months post hearing aid fitting. Neither generic fatigue scale was sensitive to changes in fatigue over time following the hearing aid fitting. In contrast, VFS-A-40 scores revealed a significant reduction in listening-related fatigue post hearing aid fitting.

In summary, the VFS-A-40 provides a reliable, valid, and sensitive measure of listening-related fatigue. However, because of its length, the measure may be better suited for research purposes rather than general clinical use. This may be especially true given that the analysis of subscale scores is warranted only in select cases (e.g., for specific research questions). For a measure to be clinically useful it must be sensitive, reliable, and user-friendly for the patient and clinician. The purpose of this study was to develop and validate a brief version of the adult VFS that would be appropriate for clinical use.

METHODS

Overview

Scale development and validation were divided into two phases. In Phase 1 we used existing VFS-A-40 survey response data (*N*=580; from Hornsby et al., 2021) and item response theory (IRT) to identify a small set of high-information items for use in a brief, 10-item version of the VFS (the VFS-A-10) that may be more appropriate for clinical usage. In Phase 2, we confirmed the quality of the selected items, and examined the dimensionality, reliability, and validity of the VFS-A-10, using data from a separate sample of 607 adults with and without hearing loss. A more detailed description of participants and procedures of each phase in the process is provided below.

Participants and Data Collection

In Phase 1, study participants were recruited locally from the patients seen at Vanderbilt Bill Wilkerson Center (VBWC) Audiology clinics and individuals from the surrounding community. In addition, we used a variety of online postings to recruit broadly. Online advertisements included a link to our consent form. Participants recruited locally were sent an email with the same link. See Hornsby et al. (2021) for a detailed description of Phase 1 participants. Briefly, 580 adults completed a large survey containing all VFS-A-40 test items. Respondents ranged in age from 18-88 years (mean = 50.2 years; standard deviation [SD] = 16 years) and were primarily female (~73%; ~6% did not report). Hearing status was determined based on their response to the question: "Do you have a hearing loss?" Those with a self-reported hearing loss were asked to rate their degree of impairment (ranging from mild/slight to profound; see Table 1).

Table 1 about here

Approximately 75% (n = 433) of respondents reported they had a hearing loss. A large majority (~82%) of those individuals reported their impairments were moderate, severe, or profound in nature (see Table 2). Phase 1 participants with hearing loss were also asked about their hearing device usage (e.g., hearing aid, cochlear implant, other). Most respondents with a hearing loss (~74%) reported using some type of hearing device in at least one ear.

Table 2 about here

For Phase 2 a separate sample of 607 adults were recruited. Participants first completed the newly developed VFS-A-10. Following this, to examine test validity, they also completed the fatigue and vigor subscales from the Profile of Mood States (POMS; McNair et al., 1971), the

Fatigue Assessment Scale (FAS; Michielsen et al., 2003), and depending on their age, the Hearing Handicap Inventory for the Elderly (HHIE; Ventry & Weinstein, 1982) or Adults (HHIA; Newman et al., 1990)—for those ≥65 or <65 years of age, respectively. Phase 2 participants included a small group of individuals (*n*=40) who had participated in prior research projects at the study sponsor's (Starkey, Inc.) laboratory. The remaining participants were recruited locally or via online advertisements as described above. Some participants were recruited while at the VBWC Audiology clinic following a routine appointment (*n*=175). These participants completed all measures on a tablet device in the presence of a research staff member. When available we also collected audiologic thresholds from this sample and from a portion of participants from the study sponsor's laboratory (*n*=176). The mean, better ear, pure-tone average threshold (PTA of thresholds at .5, 1, & 2 kHz) for this group was 42.8 dB HL (SD=21.7 dB HL). The majority (~63%) of hearing losses fell within the mild (22.9%; 26-40 dB HL), moderate (25.7%; 41-55 dB HL), and moderately-to-severe (14.9%; 56-70 dB HL) range. Approximately 26% of the sample had thresholds ≤25 dB HL and ~11% had severe-to-profound losses (thresholds of 71+ dB HL).

All other participants (*n*=392) completed the test measures online at their own convenience using personal devices. Phase 2 participants were adults aged 18-94 years (mean = 55 years; SD = 17.7 years). Most respondents were again female (~71%; ~1% did not report). Approximately 63% (*n*=379) of respondents reported they had a hearing loss. In terms of degree of impairment, in contrast to the Phase 1 sample, the majority of respondents (~58%) reported they had only a mild/slight, or no, impairment despite reporting a hearing loss. Interestingly, the percentage of hearing device users in the Phase 2 sample (~72%) was similar to that seen in the Phase 1 sample (~74%; see Table 2). In addition, a subset of Phase 2 participants (*n*=145) completed the VFS-A-10 a second time to assess test-retest reliability. The mean age for this sample was 50 years (range 22-81 years; SD = 13.9). This sample also consisted of predominantly female (78%) participants with approximately 38% reporting having a hearing loss. The range of self-reported hearing losses in the sample varied from mild/slight (~17%) to profound (~31%). The majority of self-reported losses were described as moderate (~33%) or severe (~19%) in nature. Of those with hearing loss ~65% reported using a hearing device. All data were collected using study procedures that were reviewed and approved by our local Institutional Review Board and all participants provided informed consent prior to completing any study procedures.

Phase 1 Methods: VFS-A-10 Item Selection

As we did when selecting items for the original VFS-A-40 (Hornsby et al., 2021), we used results from IRT analyses to select multiple, high information items to evaluate for inclusion in a brief version of the VFS-A. High information items are effective at identifying an individual with a given magnitude of the underlying latent construct (θ). In this case, the underlying latent construct is listening-related fatigue. *Test information* is simply the sum of item information across a set of items, also as a function of θ. High test information implies good measurement fidelity, indicating a test can provide a precise estimate of the individual's underlying latent construct (listening-related fatigue, in our case). Note that item, and test, information can vary based on the severity of the construct (e.g., mild versus severe listening-related fatigue). For example, some test items (and tests) are better able to differentiate between individuals with subtle variations in mild fatigue while others are more sensitive to variations between individuals with more severe fatigue.

We evaluated the test information provided by various combinations of items to identify the final pool of 10 items for the VFS-A-10. We purposely selected *items* such that *test* information was high over a wide range of fatigue severities, but with a focus on items that were more sensitive to detecting differences across those with severe fatigue. Consistent with our prior studies (Hornsby et al., 2021, 2022), our goal was to select items such that test information was ≥ 11.11 (test information = 1/[standard error of an IRT score^2]) over a wide range of severities of listening-related fatigue (i.e., a range of θ's). An IRT score's standard error of 0.3 was used to calculate the target test information of 11.11. This corresponds to a reliability coefficient of 0.95, a value considered acceptable in the development of other clinical scales (e.g., Cole et al., 2012; Hospers et al., 2016, Hornsby et al., 2021, 2022). The result of this iterative analysis process was the VFS-A-10.

Phase 2 Methods: Characterization and Validation of the VFS-A-10

In Phase 2 we employed a separate sample of adults (*N*=607) to characterize and validate the newly created VFS-A-10. Specifically, to characterize the tool we examined test dimensionality, item information, and item quality. To validate the VFS-A-10 we first examined test reliability using IRT and examined test-retest reliability in a subset of participants. We also examined item bias using IRT differential item functioning (DIF) analysis, assessed convergent construct validity, and examined VFS-A-10 sensitivity to differences in groups thought to vary in their level of listening-related fatigue (hearing loss versus no hearing loss). Details of the methods and results are provided below.

Phase 2.1 methods: Item and test characterization- EFA and IRT. Using data from Phase 2 participants, we conducted an EFA to explore the latent structure of our 10-item pool and completed IRT analyses to examine the characteristics of the test items. The analysis methods mirrored the approach used in our original paper (see Hornsby et al., 2021, for details) and will only be described briefly here.

EFA Methods. In Phase 2 we first investigated the latent structure (i.e., the number of dimensions and factor loading patterns) of the VFS-A-10 items by conducting a series of EFA using Mplus Version 8.3 (Muthen & Muthen, 1988-2019). An EFA was conducted using polychoric correlations (specifically, weighted least square with adjusted means and variance [WLSMV] with oblimin rotation and oblique type). We used the following empirically supported guidelines to assess the goodness of our model fit: the root-mean-square error of approximation index (RMSEA; Steiger & Lind, 1980) is less than .06, the root-mean-square residual (RMSR) is less than .08, the comparative fit index (CFI; Bentler, 1990) and Tucker-Lewis index (TTL; Tucker & Lewis, 1973) are larger than .95 (Hu & Bentler, 1999; Yu, 2002).

IRT methods. We utilized IRT analyses of Phase 2 VFS-A-10 response data to evaluate individual item and test quality. IRT analyses were conducted using a (marginal) maximum likelihood estimation method in Mplus. Based on the EFA results, a unidimensional graded response model (GRM; Samejima, 1969) was used to investigate the item properties. The GRM is a type of item response model for ordered polytomous responses and has two kinds of item parameters that are useful in determining item quality—an item discrimination parameter and item threshold parameters (see Hornsby et al., 2021, for details). Item fit was assessed using a generalized $\chi 2$ test (e.g., Kang & Chen, 2008).

Phase 2.2 methods: Reliability and validity analysis.

IRT reliability. We examined test information, using the Phase 2 data set, to confirm the VFS-A-10 was reliable (high information) and sensitive (test information ≥ 11.11) over a wide range of fatigue severities.

Test-retest reliability. To examine the stability of VFS-A-10 ratings we evaluated test-retest reliability in a subset of Phase 2 participants (*n*=145) who completed the VFS-A-10 at two different time points. The median time between tests was 28 days (range 5-90 days) with 87% of respondents completing the second test within 14-35 days of the first test. Several approaches were used to assess temporal stability. A Wilcoxon signed ranks test was used to compare mean differences in initial (T1) and retest (T2) scores. In addition, we examined associations between individual T1 and T2 scores using Spearman's correlation (r_s) coefficients (i.e., test–retest reliability coefficients) and intraclass correlation coefficients (ICCs).

IRT DIF analyses. We conducted IRT DIF analyses to confirm that the VFS-A-10 test items measured listening-related fatigue equivalently across distinct subgroups. Specifically, we assessed DIF as a function of age (18-64 years old versus ≥ 65 years old), gender (male versus female), and self-reported hearing loss (yes versus no). DIF analyses were implemented using lordif package (Choi, Gibbons, & Crane, 2011) in R (R Core Team). An ordinal logistic regression model in conjunction with IRT scale scores as a matching criterion was chosen to detect DIF items. For each item, DIF was evaluated assuming a uniform effect (the effect is constant across trait levels) and non-uniform effect (the effect varies across trait levels). DIF items were detected using likelihood ratio test to compare specific ordinal logistic regression models.[2] For each comparison, twice the difference in log likelihoods was compared to

[2] The DIF detection was made by comparing three nested ordinal logistic regression models: (a) **Model 1**: the cumulative probability that the actual item response falls in category *k* or higher =

a χ^2 distribution with a specified *df*. Type I error rate, alpha=0.01, was chosen. In addition, the pseudo R^2 measure was used to quantify the DIF effect size. Zumbo (1999) suggests classifying DIF based on the pseudo R^2 statistic as negligible (< 0.13), moderate (between 0.13 and 0.26), and large (> 0.26).

Convergent construct validity. We assessed convergent construct validity by examining associations between VFS-A-10 scores and scores obtained from several related, but distinct, measures including two generic measures of fatigue and/or vigor (i.e., the POMS fatigue and vigor subscales and the FAS) and a measure of perceived hearing difficulties, the HHIE/A. The full POMS is a 65-item scale that is designed to assess several mood states, including fatigue and vigor. The POMS is a generic measure that has been used extensively in diverse populations, including hearing loss (Hornsby & Kipp, 2016) and has good psychometric properties (i.e., internal reliability, content and construct validity, and test–retest reliability; McNair & Heuchert 2010). In this study respondents completed only the fatigue and vigor subscales of the POMS. The FAS is a generic, unidimensional, measure of fatigue that has also been widely used in diverse populations, including hearing loss (Alhanbali et al., 2017; Holman, Drummond & Naylor, 2021). The FAS has good internal reliability and content validity (Michielsen et al., 2003; Michielsen et al., 2004). The HHIE/A is a widely used tool for assessing the social and emotional consequences of hearing loss. It has good internal consistency and test-retest

---

intercept + slope1 * latent variable, (b) **Model 2**: the cumulative probability that the actual item response falls in category *k* or higher = intercept + slope1 * latent variable + slope2 * group, and (c) **Model 3**: the cumulative probability that the actual item response falls in category *k* or higher = intercept + slope1 * latent variable + slope2 * group + slope3 * latent variable * group. *Uniform DIF* was tested by comparing the log likelihood values for Models 1 and 2 (one degree of freedom, or *df* = 1) and *non-uniform DIF* by comparing Models 2 and 3 (*df* = 1). A *total DIF* effect was evaluated by comparing Models 1 and 3 (*df* = 2).

reliability (Newman et al., 1990; Ventry & Weinstein, 1982; Weinstein, Spitzer, & Ventry, 1986).

Known-groups validity. We examined known-groups validity (Davidson, 2014), a process whereby a test is administered to groups that are known, or expected, to differ on the test outcome. The VFS-A-10 was designed to be sensitive to fatigue associated with hearing loss-related listening difficulties, therefore we compared responses between Phase 2 respondents with (n = 379) and without (n = 220) self-reported hearing loss. Given the well-known variability between self-reported hearing difficulties and measured thresholds, we also conducted exploratory analyses to examine the sensitivity of the VFS-A-10 to variations in fatigue as a function of 1) pure-tone average (PTA) threshold and 2) self-reported degree of hearing impairment (see Table 1). Prior work using generic fatigue scales has shown no association between PTA and fatigue ratings but strong associations between subjective measures of hearing difficulties and fatigue ratings (e.g., Hornsby & Kipp, 2016). We hypothesized the VFS-A-10 may be more sensitive than generic measures to fatigue associated with variations in hearing loss-related listening difficulties and degree of hearing loss.

Mann-Whitney U tests were used to compare VFS-A-10 scores between groups (no hearing loss vs. hearing loss and variations in self-reported degree of hearing impairment from mild to profound, Table 1). Effect sizes of r = .1, .3, and .5 reflect small, medium, and large effects, respectively (Fritz, Morris & Richler, 2012). Spearman’s correlation (r_s) coefficients were used to examine associations between measures of hearing difficulties (PTA and self-report) and fatigue ratings. Spearman’s method was applied because we entered the fatigue scale summed scores, as would be utilized in clinical practice. The strength of the absolute correlations was rated as weak (< 0.4), moderate (0.4 to < 0.7), or strong (≥ 0.7; Akoglu, 2018). Given the exploratory nature of the analyses, various measures of PTA, including better and worse ear averages at .5, 1, & 2 kHz, .5, 1, 2 & 4 kHz, and 1, 2 & 4 kHz, were utilized in the analyses and no corrections were made for Type 1 error rate. Depending on the measure of PTA, threshold measures were available for analysis from 173-176 participants.

RESULTS

Phase 1 results: VFS-A-10 item selection. Previous work has shown that all VFS-A-40 items, regardless of their subscale assignment, were sensitive and reliable (Hornsby et al., 2021). Therefore, items for the brief scale were selected primarily on the range over which their information was high. No attempt was made to select items equally from the various subscales. A total of ten, high-information items (1, 2, 3, and 4 items from the emotional, cognitive, social, and physical subscales, respectively) were selected to create the VFS-A-10 (see Table 3). Participant responses to all items are made using a 5-point (scored as 0-4) Likert frequency scale with response options including: Never/Almost Never (0), Rarely (1), Sometimes (2), Often (3), and Always/Almost Always (4). These are the same response anchors used with these items in the VFS-A-40.

Table 3 about here

Phase 2 results: Characterization and validation of the VFS-A-10

Phase 2.1 results: Item and test characterization- EFA and IRT.

EFA results. Results from the EFA were consistent with a unidimensional model, with all VFS-A-10 items loading highly on a single factor. Standardized factor loadings for individual items ranged from 0.81 to 0.94. In addition, a one-factor (unidimensional) model provided a good fit to our data according to most of our fit indices. Only the RMSEA value (0.106) did not meet our fit criterion of <.06. In contrast, RMSR (0.030), CFI (0.993), and TLI (0.991) values all suggested a one-factor solution provided a good model fit.

IRT results. Based on the EFA results, a unidimensional GRM was fit to the VFS-A-10 Phase 2 data. Item parameter estimates of the GRM (i.e., item discrimination and item thresholds) were examined to confirm their high quality. Item discrimination estimates were all very high in magnitude (Baker, 2001), ranging from 2.26 to 4.95 (mean = 3.57 on the logit scale) and all item thresholds were in order and well separated. Item parameter discrimination and threshold estimates of the GRM for the VFS-A-10 items are shown in Table 4. Results of the generalized $\chi 2$ test suggest all items fit the Phase 2 data well.

Insert Table 4 about here

Phase 2.2 results: Reliability and validity.

IRT reliability. The solid line in Figure 1 shows the VFS-A-10 test information curve (TIC; left axis). TICs show test information as a function of θ (theta; the magnitude of an individual's listening-related fatigue in this case). Test information is simply the sum of item information as a function of θ. Thus, it is expected that test information will decrease as the number of items in the test are decreased. Despite this, the high and broad TIC suggests the VFS-A-10 has good measurement fidelity over a wide range of fatigue severities. Specifically, VFS-A-10 test information was ≥ 11.11 for θs ranging between approximately -1.5 to 2.4. Test information was reduced for individuals with very mild ($\theta < -1.5$) or very severe ($\theta > 2.4$) listening-related fatigue. However, ~93% of individual scores from Phase 2 participants fell within the range of -1.5 to 2.4 and ~99.5% of scores were ≤ 2.4 (see Figure 1), suggesting the test provided a precise estimate of listening related fatigue for most participants in this sample.

Figure 1

Test-retest reliability. Reliability analyses were conducted using the full test-retest data set (N=145) as well as using only data from those with (n = 55) and without (n = 90) self-reported hearing loss. Multiple Wilcoxon signed ranks tests revealed no significant differences in VFS-A-10 mean summed scores obtained at baseline (T1) and approximately one month later (T2), with the mean difference in summed scores being < 1 point for all comparisons. Likewise, analyses of correlations between individual scores, using Spearman's Rho and ICC's, revealed strong, positive, and statistically significant associations between scores obtained at T1 and T2, again across all comparisons. Together these findings suggest the VFS-A-10 has good temporal stability. Results of all test-retest reliability analyses are shown in Table 5.

Table 5 about here

IRT DIF results. We conducted IRT DIF analyses to confirm that the VFS-A-10 items measured listening-related fatigue equivalently in populations that varied by age, gender, and self-reported hearing loss. Three items were flagged as DIF items based on the three, likelihood ratio χ2 statistics- two items for age (Items 8 and 10), one for gender (Item 8), and three for hearing loss (Items 1, 8, and 10), respectively. However, the DIF effect sizes (Mc-Fadden's pseudo R^2 measure) for all these items were negligible (all effect sizes were <0.13; range: 0.003 to 0.0234), suggesting all VFS-A-10 items function similarly across these distinct subgroups.

Convergent construct validity. We assessed convergent construct validity by examining associations (Spearman's correlation, r_s) between VFS-A-10 summed scores (and IRT scale scores) and two generic measures of fatigue, one vigor measure, and a measure of perceived hearing difficulties. The pattern of results was, again, the same for both measures so only results from the summed score analyses are shown. As expected, we saw weak-to-moderate associations between VFS-A-10 scores and generic measures of fatigue and vigor (POMS and FAS) and strong associations with measures of perceived hearing difficulties (HHIE/A scores; see Table 6).

Table 6 about here

Known-groups validity. Analyses were conducted separately using VFS-A-10 IRT scale scores and summed scores. The pattern of results was the same regardless of the measurement method so, for clarity, only data for the summed score analyses are reported. First, consistent with our hypothesis, results of a Mann-Whitney U test showed that respondents with self-reported hearing loss reported significantly more listening-related fatigue (*Median (Mdn)* = 18; n = 379) than those without self-reported hearing loss (*Mdn* = 5; n =220; U = 14340, z = -13.1, p <.0001, r = .55). A Mann-Whitney U statistic was applied to conservatively test for differences using the rank scores. Of note, this same between-groups comparison was conducted using POMS fatigue and vigor scores and FAS scores. These results also revealed significant differences, however, effect sizes for these comparisons were substantially reduced compared to the VFS-A-10 (r values ranged from .10 to .19). Figure 2 shows the distributions of VFS-A-10 summed scores for Phase 2 participants with and without self-reported hearing loss.

Figure 2 about here.

Second, we used a series of Mann-Whitney U tests to compare VFS-A-10 scores between groups who varied in terms of self-reported degree of hearing impairment. Results revealed a systematic increase in listening-related fatigue as self-reported degree of hearing impairment increased from none (no hearing loss; *Mdn* = 5; n = 220) to mild/slight (*Mdn* = 11; n = 85; U =

6425, $z = -4.4$, $p <.0001$, $r = .25$), to moderate ($Mdn = 17$; $n = 138$; $U = 3451$, $z = -5.2$, $p <.0001$, $r = .35$), and to a severe impairment ($Mdn = 26$; $n = 69$; $U = 3222$, $z = -3.8$, $p <.0001$, $r = .26$). However, fatigue ratings of those with severe and profound impairment ($Mdn = 23.5$; $n = 76$; $U = 2481$, $z = -.56$, $p = .576$, $r = .05$) were not significantly different. Again, of note, we replicated these analyses using FAS and POMS fatigue and vigor scores and found only a single significant difference across all analyses. POMS fatigue scores for those with a moderate self-reported impairment ($Mdn = 7$; $n = 137$) were slightly, but significantly, higher than those reporting only a slight/mild impairment ($Mdn = 7$; $n = 83$; $U = 4780$, $z = -1.98$, $p = .048$, $r = .13$). None of the other between group comparisons using the FAS or POMS (fatigue or vigor subscales) reached statistical significance (all p values were > .05). The distribution of VFS-A-10, FAS, and POMS fatigue and vigor scores as a function of impairment group are shown as boxplots in Figure 3.

Figure 3 about here.

We also examined associations between individual VFS-A-10 scores and their *self-reported degree of impairment*. Results revealed a moderate association ($r_s = .63$, $p <.0001$), consistent with prior research (e.g., Hornsby & Kipp, 2016) and the mean analyses shown in Figure 3. In contrast to prior work, we also found significant associations between the various behavioral threshold measures (i.e., PTAs) and VFS-A-10 fatigue scores, however, the strength of those associations was relatively weak. Spearman's correlations (r_s) between PTA and VFS-A-10 scores varied based on the specific comparison, but all ranged from .23 to .32. The same analyses examining associations between PTAs and fatigue or vigor scores from generic measures (i.e., POMS and FAS) revealed no significant associations (r_s values ranged from -0.12 to 0.06, all p values >.05). We also examined associations between these generic measures (POMS and FAS) and *self-reported degree* of impairment. Results revealed significant correlations for all comparisons (all p values were <.05); however, in contrast to the VFS-A-10 findings, the associations were relatively weak (r_s values ranged from -0.13 to 0.23).

SUMMARY AND DISCUSSION

In this paper we describe the development and validation of a brief, 10-item, version of the Vanderbilt Fatigue Scale for Adults-- the VFS-A-10. The VFS-A-10 is a unidimensional measure designed to assess listening-related fatigue. Given the increased incidence of fatigue among adults and children with hearing loss and its negative effects on quality of life, we believe the VFS-A-10 will provide an important tool for clinicians and researchers. Such a reliable and valid tool is especially important given the weak association between listening-related fatigue and audiometric measures. Scale items are shown in Table 3. An electronic copy of the full VFS-A-10, along with information and electronic copies of all other versions of the VFS, can be freely downloaded from the VFS website; https://www.vumc.org/vfs/.

Multiple analyses confirmed the VFS-A-10 is a sensitive and reliable measure of listening-related fatigue. IRT analyses revealed the VFS-A-10 had good fidelity (test information ≥11.11) over a wide range of severity levels (see Figure 1). In fact, VFS-A-10 test information was ≥11.11 for θ values up to 2.4 (a very high level of listening-related fatigue). This is similar to, or better than, the fidelity of the full VFS-A-40 and its subscales (Hornsby et al., 2021) and the VFS-Peds (Hornsby et al., 2022). In addition, an analysis of test-retest reliability using data from a subset of Phase 2 respondents found the VFS-A-10 had excellent (e.g., Speaman's rho = .88) temporal stability over an approximately one-month period. Note that this association remained essentially unchanged (Spearman's rho = .89) even when limiting the analysis to re-test scores collected within 14-35 days of the original (n = 125). For comparison, Spearman's correlation coefficients ranged from .60-.69 for the VFS-A-40 test-retest analyses (Hornsby et al., 2021). However, these values reflect comparisons over an approximately 3-month period compared to the 1-month period for the VFS-A-10 analyses. Given the longer time frame and the understanding that mood states, such as fatigue, are expected to vary over time, the weaker VFS-A-40 correlations are not unexpected.

The VFS-A-10 was also shown to be highly sensitive to variations in listening-related fatigue between individuals with and without self-reported (see Figure 2), and objectively measured (see known-groups validity section), hearing loss. Prior work using generic fatigue and vigor measures have also observed significant differences between these groups (Alhanbali et al., 2017; Hornsby & Kipp, 2016; Hornsby et al., 2017; Sindhar et al., 2021), however, this is not a universal finding (e.g., Hornsby & Kipp, 2016; Dwyer et al., 2019). In addition, the magnitude of hearing loss effects can vary widely across studies (see Hornsby et al., 2017 and Hornsby et al., 2021, for reviews). This study adds to the literature by highlighting the increased sensitivity of the VFS-A-10, compared to generic instruments such as the POMS and FAS, to fatigue experienced by individuals with self-reported hearing loss (see Phase 2.2 Results: Known-groups validity section).

In contrast to the mean effects of hearing loss noted above, counter-intuitively but uniformly, studies in adults and children have shown no association between PTA and ratings from generic fatigue measures (Alhanbali et al., 2017; Hornsby & Kipp, 2016; Hornsby et al., 2017). In the current study we asked respondents with self-reported hearing loss to qualitatively describe their degree of hearing impairment using response options ranging from mild/slight to profound (see Table 1); and, in a subset of respondents, we also obtained audiometric pure-tone thresholds from clinical records. These data provided us with the opportunity to examine the sensitivity of the VFS-A-10 to variations in listening-related fatigue associated with subjective and objective measures of hearing impairment. VFS-A-10 scores were significantly higher for those with self-reported hearing loss and scores increased systematically as the magnitude of self-reported hearing difficulties increased (at least up to severe losses). In addition, when using the VFS-A-10 we observed a significant, albeit weak, correlation between PTA measures and fatigue ratings. However, consistent with prior research, when we replicated these analyses using generic fatigue and vigor data collected in this study no significant correlations were found. This finding is consistent with prior work (e.g., Dwyer et al., 2019; Holman et al., 2021) thus confirming the limited sensitivity of generic measures to fatigue effects associated with variations in hearing thresholds and listening difficulties.

In terms of clinical use, the VFS-A-10 is a unidimensional measure that quantifies long-term listening-related fatigue (i.e., the fatigue experienced over the past week) in adults. The scale is simple and easy to use. The full VFS-A-10, including instructions, has a Flesch-Kincaid grade reading level of 6.5. The mean/median grade reading level for individual items is 6.8/6.7 (SD= 1.3; range = 5.8-8.8). Importantly, it has been our experience that most adults are able to complete the scale on their own in 2-3 minutes. Supporting this anecdotal finding, an analysis of missing data from our Phase 2 sample (*N*=607) showed most individuals (~96%) completed all VFS-A-10 test items, with only ~3.5% of participants missing a single item. Only 1 person failed to respond to more than 2 items.

Like the other versions of the VFS, the VFS-A-10 can be scored by simply summing the item responses. When interpreting scores, it is helpful to keep in mind that the scale uses a 5-point Likert response format to ask about the frequency of listening-related fatigue problems being experienced in a typical week. Response options range from 0, meaning the respondent never has problems, to 4, meaning they almost always have problems. Scores of 1, 2 or 3 indicate a specific issue occurs rarely, sometimes, or often, respectively. Given the scale consists of 10 items, summed scores ≤10 mean the respondent experiences problems with listening-related fatigue, but the problems are relatively rare or, with a score of 0, they never or almost never occurred. In contrast, a score ≥30 means this individual is experiencing listening-related fatigue issues often or almost always.

Figure 2 shows the distribution of scores from individuals in our Phase 2 sample and shows that <u>all</u> adults who reported no hearing loss had VFS-A-10 scores ≤ 26. In fact, the majority (~53%) had VFS-A-10 scores ≤ 5 and ~88% had scores ≤ 15 (i.e., suggesting that fatigue-related problems were occurring only rarely to sometimes). Note that these percentages are based on the actual number of sample participants scoring in the designated ranges and are not percentile ranks based on extrapolation from a normal distribution.

Although these results are predicated on a convenience sample and require further replication, they do suggest that for most individuals without a self-reported hearing loss, problems with listening-related fatigue are relatively infrequent. In contrast, Phase 2 respondents with self-reported hearing loss were much more likely to report frequent problems. Approximately 25% had scores > 26 and ~13% had scores ≥ 30, suggesting their problems were occurring often to always. Future research is needed to better understand the functional impact of these frequent problems.

In addition to summed scoring, VFS-A-10 item responses can also be used to calculate IRT scale scores, although this approach requires knowledge of IRT analysis methods and the use of statistical software packages capable of such analyses. While summed scoring may be more appropriate for quick clinical use, because IRT scoring provides a more precise estimate of an individual's listening-related fatigue it may be useful for research purposes. For those interested, we have developed R code that can be used to calculate IRT scale scores based on VFS-A-10 item responses. The code can be downloaded at the following link: https://www.vumc.org/vfs/

In summary, listening-related fatigue is an important issue for many adults with hearing difficulties. The VFS-A-10 allows for the quick and reliable assessment of listening-related fatigue in this population. We believe the tool will be useful for identifying those individuals most affected by listening-related fatigue and for assessing the effectiveness of any interventions designed to reduce their fatigue. For example, prior work using the full VFS-A-40 has shown that well-fit hearing aids can reduce listening-related fatigue in new hearing aid recipients (Holman, Drummond, & Naylor, 2021). While designed with a focus on the fatigue experienced by adults with hearing loss, we believe the VFS-A-10 could also be useful as a screening instrument for other populations who struggle with listening and understanding—for example non-native speakers, individuals with auditory processing deficits, and/or individuals with hidden hearing loss. However, additional research is required to determine the prevalence and magnitude of such difficulties in these populations.

REFERENCES

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91-93. https://doi.org/10.1016/j.tjem.2018.08.001

Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear and Hearing*, *38*(1), e39-e48. https://doi.org/10.1097/aud.0000000000000361

Amato, M., Ponziani, G., Rossi, F., Liedl, C., Stefanile, C. & Rossi, L. (2001). Quality of Life in Multiple Sclerosis: The Impact of Depression, Fatigue and Disability. *Multiple Sclerosis*, *7*(5), 340–344. doi:10.1177/135245850100700511

Baker, F. B. (2001). *The basics of item response theory* (2nd. Ed.). College Park: ERIC Clearinghouse on Assessment and Evaluation, University of Maryland. Retrieved from https://eric.ed.gov/?id=ED458219.

Bentler, P. M. (1990). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *107*, 238-246. http://dx.doi.org/10.1037/0033-2909.107.2.238

Bess, F. H., & Hornsby, B. W. (2014). Commentary: listening can be exhausting--fatigue in children and adults with hearing loss. *Ear and Hearing*, *35*(6), 592-599. https://doi.org/10.1097/AUD.0000000000000099

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, *39*(8), 1. https://doi.org/10.18637/jss.v039.i08

Cole, D. A., Cho, S.-J., Martin, N. C., Youngstrom, E. A., March, J. S., Findling, R. L., Compas, B. E., Goodyer, I. M., Rohde, P., Weissman, M., Essex, M. J., Hyde, J. S., Curry, J. F., Forehand, R., Slattery, M. J., Felton, J. W., & Maxwell, M. A. (2012). Are increased weight and appetite useful indicators of depression in children and adolescents? *Journal of Abnormal Psychology*, *121*(4), 838-851. https://doi.org/10.1037/a0028175

Davidson, M. (2014). Known-Groups Validity. In: Michalos, A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_1581

Davis, H., Schlundt, D., Bonnet, K., Camarata, S., Bess, F. H., & Hornsby, B. W. (2021a). Understanding listening-related fatigue: Perspectives of adults with hearing loss. *International Journal of Audiology*, *60*(6), 458-468. https://doi.org/10.1080/14992027.2020.1834631

Davis, H., Schlundt, D., Bonnet, K., Hornsby, B. W., & Bess, F. H. (2021b). Listen-Related Fatigue in Children with Hearing Loss: Perspectives of Children, Parents, and School

Professionals. *American Journal of Audiology, 30(*4), 929-940. https://doi.org/10.1044/2021_AJA-20-00216

Dwyer, R. T., Gifford, R. H., Bess, F. H., Dorman, M., Spahr, A., & Hornsby, B. W. Y. (2019). Diurnal Cortisol Levels and Subjective Ratings of Effort and Fatigue in Adult Cochlear Implant Users: A Pilot Study. *American Journal of Audiology*, *28*(3), 686-696. https://doi.org/10.1044/2019_AJA-19-0009

Dittner, A. J., Wessely, S. C., & Brown, R. G. (2004). The assessment of fatigue: a practical guide for clinicians and researchers. *Journal of Psychosomatic Research*, *56*(2), 157-170. https://doi.org/10.1016/S0022-3999(03)00371-4.

Eddy, L., & Cruz, M. (2007). The Relationship between Fatigue and Quality of Life in Children with Chronic Health Problems: A Systematic Review. *Journal for Specialists in Pediatric Nursing, 12*(2): 105–114. doi:10.1111/j.1744-6155.2007.00099.x.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Gaba, D. M., & Howard, S. K. (2002). Patient safety: fatigue among clinicians and the safety of patients. *New England Journal of Medicine, 347*(16), 1249-1255. https://doi.org/10.1056/NEJMsa020846

Hambleton, R.K. & Swaminathan, H. (2013). Item Response Theory: Principles and Applications. Springer Science + Business Media, LLC.

Hockenberry-Eaton, M., Hinds, P., Howard, V., Gattuso, J., O'Neill, J., Alcoser, P., Bottomley, S., Kline, N., & Euell, K. (1999). Developing a conceptual model for fatigue in children. *European Journal of Oncology Nursing, 3*(1), 5-11. https://doi.org/10.1016/1462-3889(91)80005-7

Hockey, R. (2013). *The psychology of fatigue: Work, effort and control*. Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9781139015394

Holman, J. A., Drummond, A., Hughes, S. E., & Naylor, G. (2019). Hearing impairment and daily-life fatigue: A qualitative study. *International Journal of Audiology*, *58*(7), 408-416.

Holman, J. A., Drummond, A., & Naylor, G. (2021). Hearing aids reduce daily-life fatigue and increase social activity: a longitudinal study. *Trends in hearing*, *25*, https://doi.org/10.1177/23312165211052786.

Hornsby, B. (2013). The Effects of Hearing Aid Use on Listening Effort and Mental Fatigue Associated With Sustained Speech Processing Demands. *Ear and Hearing*, *34*(5), 523-534 510.1097/AUD.1090b1013e31828003d31828008.

Hornsby, B., Camarata, S., Cho, S.-J., Davis, H., McGarrigle, R., & Bess, F. H. (2021). Development and Validation of the Vanderbilt Fatigue Scale for Adults (VFS-A). *Psychological Assessment, 33(*8), 777-788. . https://doi.org/10.1037/pas0001021

Hornsby, B., Camarata, S., Cho, S.-J., Davis, H., McGarrigle, R., & Bess, F.H. (2022). Development and Evaluation of Pediatric Versions of the Vanderbilt Fatigue Scale (VFS-Peds) for Children with Hearing Loss. *Journal of Speech, Language, and Hearing Research.*

Hornsby, B., Gustafson, S. J., Lancaster, H., Cho, S.-J., Camarata, S., & Bess, F. H. (2017). Subjective Fatigue in Children With Hearing Loss Assessed Using Self- and Parent-Proxy Report. *American Journal of Audiology, 26*(3S), 393-407. doi: 10.1044/2017_AJA-17-0007

Hornsby, B. & Kipp, A. M. (2016). Subjective Ratings of Fatigue and Vigor in Adults With Hearing Loss Are Driven by Perceived Hearing Difficulties Not Degree of Hearing Loss. *Ear and Hearing, 37*(1), e1-e10. doi: 10.1097/aud.0000000000000203

Hornsby, B., Naylor, G., & Bess, F. H. (2016). A Taxonomy of Fatigue Concepts and Their Relation to Hearing Loss. *Ear Hear, 37 Suppl 1*, 136S-144S. doi: 10.1097/AUD.0000000000000289

Hospers, J. M. B., Smits, N., Smits, C., Stam, M., Terwee, C. B., & Kramer, S. E. (2016). Reevaluation of the Amsterdam Inventory for Auditory Disability and Handicap Using Item Response Theory. *Journal of Speech, Language, and Hearing Research, 59*(2), 373-383. https://doi.org/doi:10.1044/2015_JSLHR-H-15-0156

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. http://dx.doi.org/10.1080/10705519909540118

Kang, T., & Chen, T. T. (2008). Performance of the generalized S-X2 item fit index for polytomous IRT models. *Journal of Educational Measurement, 45*(4), 391–406. https://doi.org/10.1111/j.1745-3984.2008.00071.x

McGarrigle, R., Knight, S., Rakusen, L., Geller, J., & Mattys, S. (2021a). Older adults show a more sustained pattern of effortful listening than young adults. *Psychology and aging*, *36*(4), 504–519. https://doi.org/10.1037/pag0000587

McGarrigle, R., Knight, S., Hornsby, B. W. Y., & Mattys, S. (2021b). Predictors of listening-related fatigue across the adult life span. *Psychological Science, 32*(12), 1937–1951. https://doi.org/10.1177/09567976211016410

McNair D., & Heuchert, J. (2010). *Profile of Mood States Technical Update*. Toronto, ON: Multi-Health Systems, Inc.

McNair, D., Lorr, M., & Droppleman, L. (1971). *Profile of mood states*. San Diego, Calif.: Educational and Industrial Testing Service.

Michielsen, H. J., De Vries, J., & Van Heck, G. L. (2003). Psychometric qualities of a brief self-rated fatigue measure: The Fatigue Assessment Scale. *Journal of psychosomatic research*, *54*(4), 345-352. https://doi.org/10.1016/s0022-3999(02)00392-6

Michielsen, H. J., De Vries, J., Van Heck, G. L., Van de Vijver, F. J., Sijtsma, K. (2004). Examination of the dimensionality of fatigue. European Journal of Psychological Assessment, 20(1), 39–48. https://doi.org/10.1027/1015-5759.20.1.39

Muthen, L. & Muthen, B. (1998-2017). Mplus User's Guide (8th Edition). Muthen & Muthen. Patrick, D. L., & Deyo, R. A. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Medical Care,* S217-S232.

Newman, C., Weinstein, B., Jacobson, G., & Hug, G. (1990). The Hearing Handicap Inventory for Adults: psychometric adequacy and audiometric correlates. *Ear and Hearing, 11(6),* 430-433.

Patrick, D. L., & Deyo, R. A. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Medical care*, S217-S232.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., & Mackersie, C. L. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing, 37,* 5S-27S.

Piper, B. F., Dibble, S. L., Dodd, M. J., Weiss, M. C., Slaughter, R. E., & Paul, S. M. (1998). The revised Piper Fatigue Scale: psychometric evaluation in women with breast cancer. *Oncol Nurs Forum, 25*(4), 677-684.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ravid, S., Afek, I., Suraiya, S., Shahar, E., & Pillar, G. (2009). Kindergarten children's failure to qualify for first grade could result from sleep disturbances. *Journal of Child Neurology, 24*(7), 816-822. https://doi.org/10.1177/0883073808330766

Ricci, J. A., Chee, E., Lorandeau, A. L., & Berger, J. (2007). Fatigue in the U.S. workforce: prevalence and implications for lost productive work time. *Journal Occupational Environmental Medicine, 49*(1), 1-10.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika, 34*, 1-97. http://doi.org/10.1007/BF03372160

Sindhar, S., Friesen, T. L., Carpenter, D., Kesser, B., & Lieu, J. (2021). Fatigue in children with unilateral and bilateral hearing loss. Otology & Neurotology, 42(9), e1301–e1307. https://doi.org/10.1097/MAO.0000000000003225

Steiger, J. H., & Lind, J. (1980). *Statistically-based tests for the number of common factors.* Paper presented at the Annual Spring Meeting of the Psychometric Society, IA.

Strohschein, F. J., Kelly, C. G., Clarke, A. G., Westbury, C. F., Shuaib, A., & Chan, K. M. (2003). Applicability, validity, and reliability of the Piper Fatigue Scale in postpolio patients. *American journal of physical medicine & rehabilitation*, *82*(2), 122-129. DOI: 10.1097/01.PHM.0000046632.06206.80

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1–10. https://dxdoi.org/10.1007/BF02291170

Varni, J. W., Burwinkle, T. M., Katz, E. R., Meeske, K., & Dickinson, P. (2002). The PedsQL in pediatric cancer: reliability and validity of the Pediatric Quality of Life Inventory Generic Core Scales, Multidimensional Fatigue Scale, and Cancer Module. *Cancer*, *94*(7), 2090-2106. https://doi.org/10.1002/cncr.10428

Ventry, I., & Weinstein, B. (1982). The hearing handicap inventory for the elderly: a new tool. *Ear and Hearing*, *3*(3), 128-134.

Weinstein, B. E., Spitzer, J. B., Ventry, I. M. (1986). Test-retest reliability of the hearing handicap inventory for the elderly. *Ear & Hearing*, *7*, 295–299. DOI: 10.1097/00003446-198610000-00002

Whitehead, L. (2009). The measurement of fatigue in chronic illness: a systematic review of unidimensional and multidimensional fatigue measures. *Journal of Pain & Symptom Management*, *37*(1), 107-128. https://doi.org/10.1016/j.jpainsymman.2007.08.019.

Yu, C. Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes (Unpublished doctoral dissertation). University of California, Los Angeles, CA.

Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF). *Ottawa: National Defense Headquarters,* 1-57.

Figure Legends.

Figure 1. Distribution of IRT scores in our Phase 2 sample in relation to VFS-A-10 test information. Distribution of IRT scale scores for 600 respondents (Grey bars; right y-axis represents percentage of respondent IRT scores falling within a certain range) and test information of the VFS-A-10 (solid line; left y-axis represents test information) as a function of IRT score (x-axis). The dashed line represents test information of 11.1 which corresponds to a reliability coefficient of .95.

Figure 2. Histograms showing the distribution of VFS-A-10 summed scores for Phase 2 participants with (grey bars) and without (black bars) self-reported hearing loss.

Figure 3. Box and whisker plot showing median values (solid line within grey box), 25^{th} and 75^{th} percentiles (lower and upper bounds of the grey box), 10^{th} and 90^{th} percentiles (error bars) and outlying points (filled black circles) for VFS-A-10, FAS, and POMS fatigue and vigor subscale response data as a function of self-reported hearing impairment. Solid black lines above the error bars identify significant (*) and not significant (N.S.) differences between adjacent groups.

Figure1

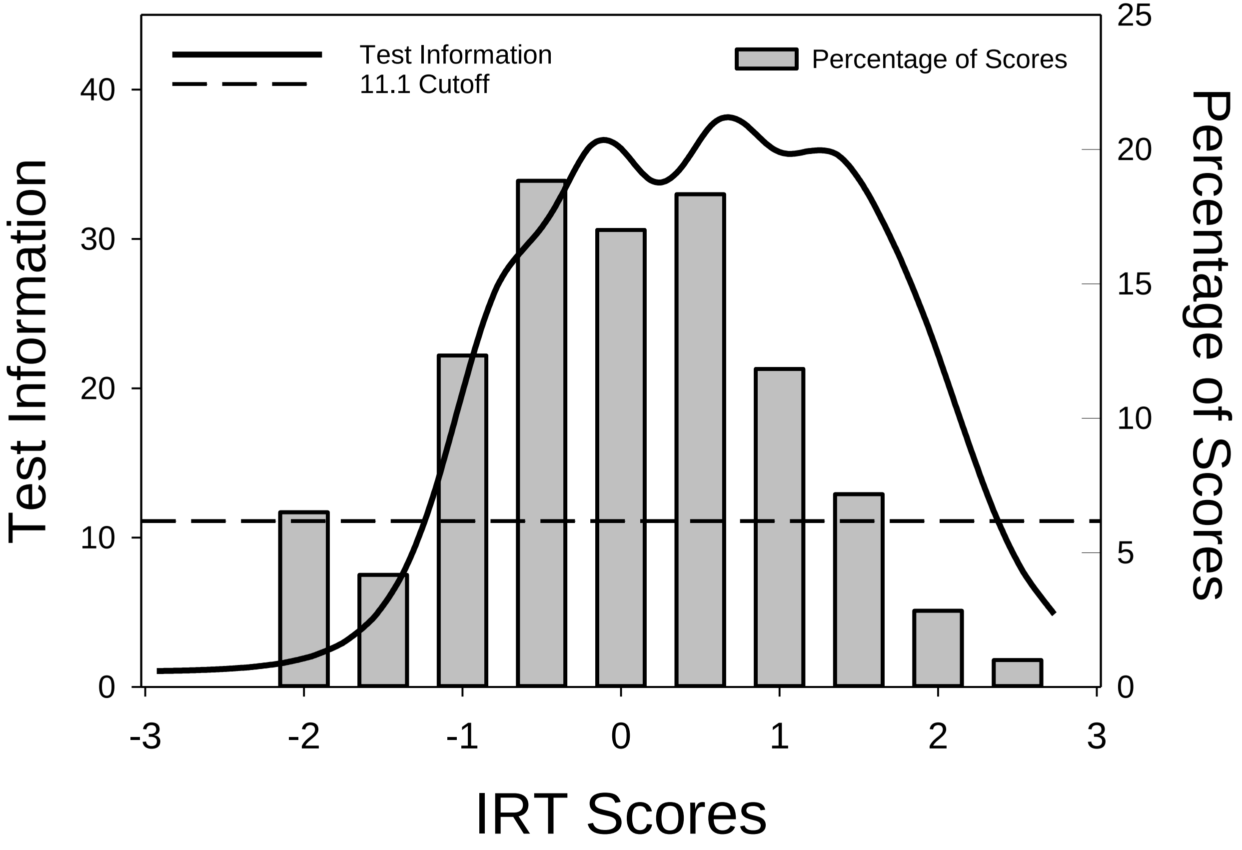
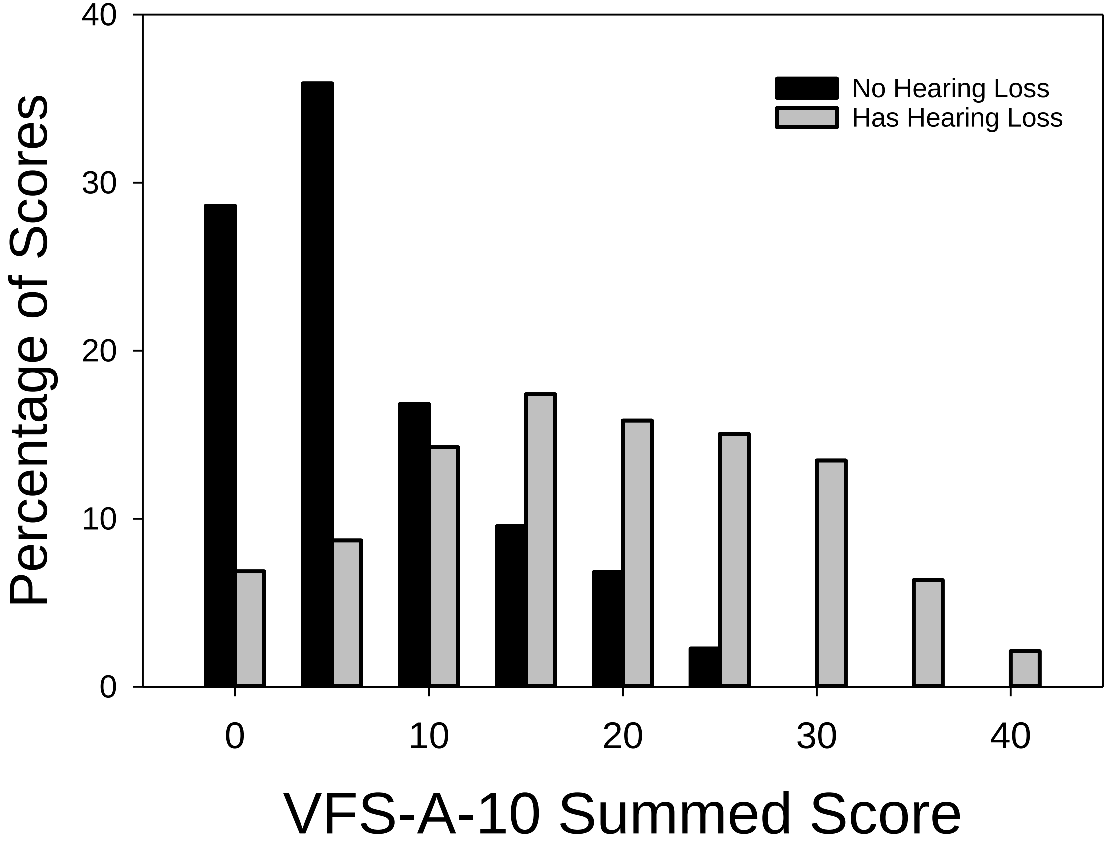
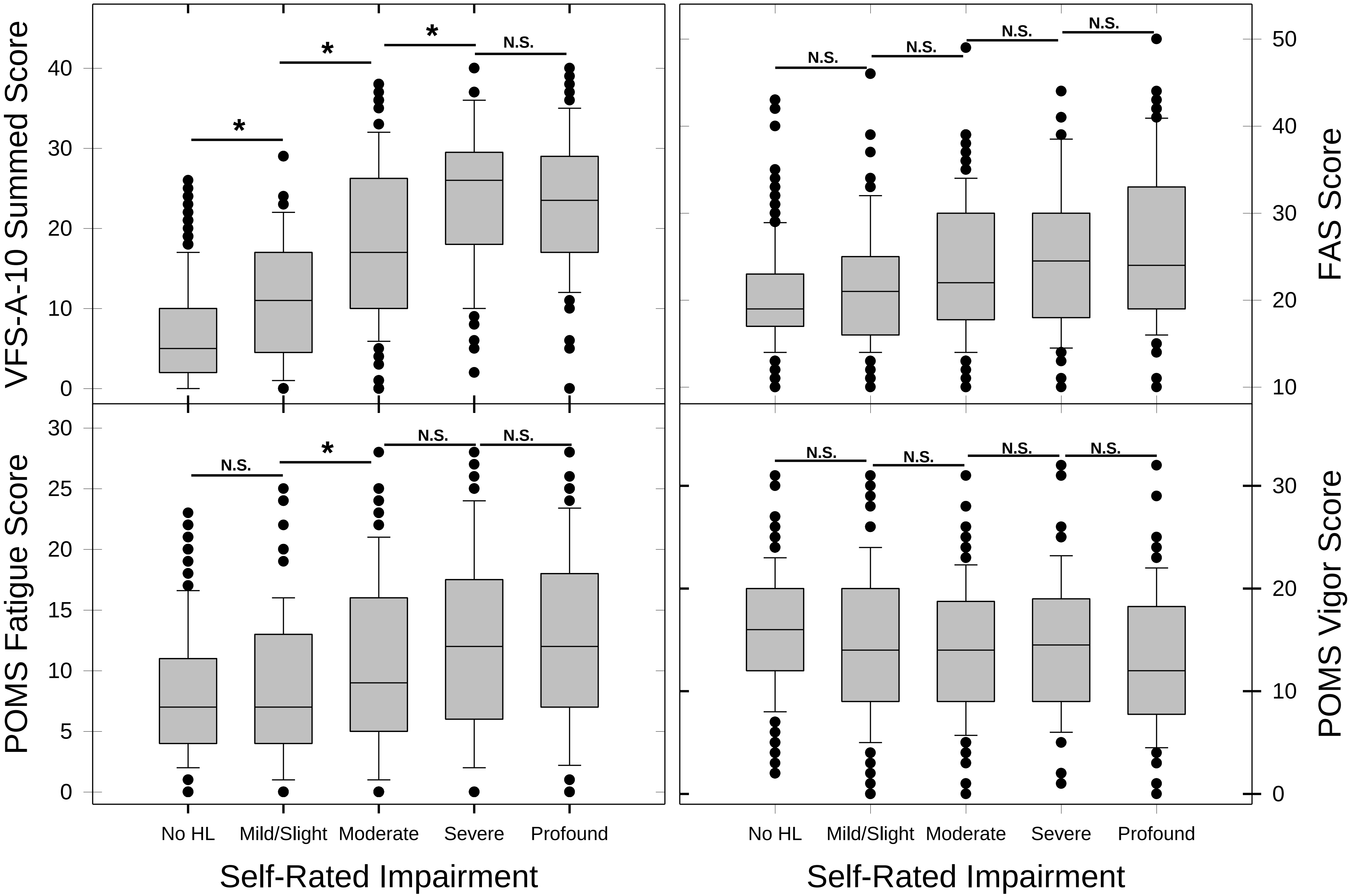

Test Information
11.1 Cutoff
Percentage of Scores
Test Information
Percentage of Scores
IRT Scores
0
10
20
30
40
0
5
10
15
20
25
-3
-2
-1
0
1
2
3

Figure2



Percentage of Scores
40
30
20
10
0
No Hearing Loss
Has Hearing Loss
0
10
20
30
40
VFS-A-10 Summed Score

Figure3



VFS-A-10 Summed Score
0
10
20
30
40
*
*
*
N.S.
FAS Score
10
20
30
40
50
N.S.
N.S.
N.S.
N.S.
POMS Fatigue Score
0
5
10
15
20
25
30
N.S.
*
N.S.
N.S.
POMS Vigor Score
0
10
20
30
N.S.
N.S.
N.S.
N.S.
No HL
Mild/Slight
Moderate
Severe
Profound
Self-Rated Impairment
No HL
Mild/Slight
Moderate
Severe
Profound
Self-Rated Impairment

TABLES

Table 1. Self-reported hearing difficulty of Phase 2 respondents. Respondents that reported they had a hearing loss were asked to rate their degree of impairment using this scale. Grading of hearing difficulty is a modification of the WHO Grades of Hearing Impairment (Mathers, Smith, & Concha, 2000).

Rating	Grade of hearing difficulty due to self-reported hearing loss
1	No Impairment: I have no or very slight hearing problems. I am able to hear whispers.
2	Mild/Slight Impairment: I am able to hear and understand speech spoken in a normal voice from a distance of about 3 feet (conversational distance). I have trouble hearing and understanding soft (whispered) speech or speech from a farther distance, especially if there is background noise.
3	Moderate Impairment: I am able to hear and understand speech spoken in a louder voice at a distance of about 3 feet (conversational distance). I have trouble hearing and understanding speech spoken at a normal level even at close distances, especially if there is background noise.
4	Severe Impairment: I am able to hear some speech without using a hearing device (hearing aid, cochlear implant), if the person is shouting or speaking in a loud voice.
5	Profound Impairment: I am unable to hear and understand even a shouted voice if I am not wearing amplification.

Table 2. Number and percentage [shown in ( )] of Phase 1 and Phase 2 respondents with self-reported hearing loss, a given grade of hearing impairment and hearing device use.

	Phase 1 Sample (*N*=580)	Phase 2 Sample (*N*=607)
Self-reported Hearing Loss (HL)		
No HL	143 (24.7%)	220 (36.2%)
Has HL	433 (74.7%)	379 (62.4%)
Did not report	4 (0.7%)	8 (1.3%)
Grade of impairment	*n* = 433	*n* = 379
No Impairment	6 (1.4%)	85 (22.2%)
Mild/Slight Impairment	71 (16.4%)	138 (36.7%)
Moderate Impairment	159 (36.7%)	69 (18.2%)
Severe Impairment	102 (23.6%)	76 (20.1%)
Profound Impairment	95 (21.9%)	0 (0%)
Did not report	0 (0%)	11 (2.9%)
Hearing device Use	*n* = 433	*n* = 379
Does not use a hearing device	105 (24.2%)	77 (20.3%)
Uses a hearing device	322 (74.4%)	275 (72.6%)
Did not report	6 (1.4%)	27 (7.1%)

Table 3. VFS-A-10 test items along with original VFS-A-40 item number and subscale domain.

VFS-A-40 Item	VFS-A-40 Subscale	VFS-A-10 Item
4	Physical	1. I feel worn out from everyday listening.
31	Physical	2. Struggling to listen and understand makes me feel tired.
11	Social	3. I get so exhausted from listening that I cannot do the things I enjoy.
15	Physical	4. I schedule my day to avoid getting tired from listening.
13	Cognitive	5. I get so tired from listening that I start to miss details in a conversation.
19	Physical	6. I get so exhausted from listening that I go to bed early.
26	Social	7. I withdraw when I am unable to follow conversations in noisy places.
6	Social	8. Feeling tired from listening causes strain on my relationships.
25	Emotional	9. I feel emotionally drained when it is hard for me to listen and understand.
3	Cognitive	10. It takes a lot of energy to listen and understand.

Table 4. *VFS-A-10 Item Parameter Discrimination and Threshold Estimates (EST) and Standard Errors (SE)*

Items	Discrimination		Threshold 1		Threshold 2		Threshold 3		Threshold 4	
	EST	SE	EST	SE	EST	SE	EST	SE	EST	SE
1	4.169	0.292	-3.515	0.306	-0.486	0.224	3.317	0.295	6.637	0.424
2	4.948	0.354	-4.22	0.347	-1.202	0.271	2.986	0.316	6.528	0.46
3	4.618	0.317	-0.644	0.246	2.5	0.289	5.755	0.415	8.869	0.624
4	2.994	0.228	0.834	0.184	2.547	0.232	4.473	0.331	6.058	0.417
5	3.705	0.252	-2.741	0.249	-0.372	0.203	2.679	0.249	6.398	0.413
6	2.742	0.196	0.289	0.167	2.053	0.194	3.898	0.258	5.83	0.362
7	2.261	0.147	-2.712	0.183	-1.426	0.155	0.883	0.147	3.348	0.209
8	3.055	0.193	-0.598	0.177	1.489	0.19	4.16	0.276	6.574	0.398
9	3.408	0.229	-2.329	0.221	-0.502	0.191	2.332	0.224	4.733	0.322
10	3.772	0.252	-3.166	0.255	-1.152	0.215	1.833	0.228	4.288	0.307

Note. Values are based on the unidimensional graded response model.

Table 6. Spearman's rank correlations (r_s) between VFS-A-10 summed scores and FAS, POMS Fatigue and Vigor, and HHIE/A Total, Emotional, and Social, scores. All correlations are significant at $p < .0001$.

	VFS-A-10	FAS	POMS Fatigue	POMS Vigor	HHIE/A Total	HHIE/A Emotional	HHIE/A Social
VFS-A-10		.565	.550	-.358	.804	.807	.792
FAS			.762	-.650	.454	.472	.431
POMS Fatigue				-.619	.406	.430	.383
POMS Vigor					-.291	-.312	-.269
HHIE/A Total						.982	.987
HHIE/A Emotional							.944
HHIE/A Social							

Table 5. Mean VFS-A-10 summed scores and standard errors [] at Time 1 and Time 2 for a subset of Phase 2 respondents (*n*=145) with (*n*=55) and without (*n*=90) hearing loss (HL). Wilcoxon Z and the resultant p-values in () are shown in the fourth column. Spearman's Rho correlation (r_s) and Intraclass correlation coefficient (ICC) values for comparisons of scores at T1 and T2 are shown in the fifth and final columns, respectively. **Bolded** values are significant at the 0.01 level (2-tailed). ICC 95% confidence intervals are shown in {}.

Test Group	Time 1	Time 2	Wilcoxon Z	Spearman's Rho	ICC
All (No HL & HL groups) Groups	11.7 [0.84]	12.0 [0.83]	-.893 (.372)	**.88 (<0.001)**	**.95 {.93-.97}**
No HL Group	6.5 [0.64]	6.9 [0.68]	-.703 (.482)	**.77 (<0.001)**	**.87 {.80-.92}**
HL Group	20.1 [0.84]	20.4 [0.83]	-.357 (.721)	**.87 (<0.001)**	**.94 {.90-.97}**