Reevaluation of the Amsterdam Inventory for Auditory Disability and Handicap Using Item Response Theory

Boeschen Hospers, J. M.; Smits, N.; Smits, C.; Stam, M.; Terwee, C.B.; Kramer, S.E.

DOI
10.1044/2015_JSLHR-H-15-0156

Publication date
2016

Document Version
Final published version

Published in
Journal of Speech, Language, and Hearing Research

Citation for published version (APA):
Boeschen Hospers, J. M., 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*, 373-383. https://doi.org/10.1044/2015_JSLHR-H-15-0156

JSLHR

Research Article

Reevaluation of the Amsterdam Inventory for Auditory Disability and Handicap Using Item Response Theory

J. Mirjam Boeschen Hospers,[a] Niels Smits,[b] Cas Smits,[a] Mariska Stam,[a] Caroline B. Terwee,[c] and Sophia E. Kramer[a]

Purpose: We reevaluated the psychometric properties of the Amsterdam Inventory for Auditory Disability and Handicap (AIADH; Kramer, Kapteyn, Festen, & Tobi, 1995) using item response theory. Item response theory describes item functioning along an ability continuum.
Method: Cross-sectional data from 2,352 adults with and without hearing impairment, ages 18–70 years, were analyzed. They completed the AIADH in the web-based prospective cohort study "Netherlands Longitudinal Study on Hearing." A graded response model was fitted to the AIADH data. Category response curves, item information curves, and the standard error as a function of self-reported hearing ability were plotted.
Results: The graded response model showed a good fit. Item information curves were most reliable for adults who reported having hearing disability and less reliable for adults with normal hearing. The standard error plot showed that self-reported hearing ability is most reliably measured for adults reporting mild up to moderate hearing disability.
Conclusions: This is one of the few item response theory studies on audiological self-reports. All AIADH items could be hierarchically placed on the self-reported hearing ability continuum, meaning they measure the same construct. This provides a promising basis for developing a clinically useful computerized adaptive test, where item selection adapts to the hearing ability of individuals, resulting in efficient assessment of hearing disability.

Hearing impairment is a chronic condition that may lead to limitations in daily communication and emotional and social loneliness (Granberg et al., 2014; Pronk et al., 2011). The degree of hearing loss is usually defined by the average pure-tone threshold (Ramkissoon & Cole, 2011). However, several studies have shown that pure-tone measures alone provide an insufficient reflection of real-world hearing difficulties (Diao et al., 2014; Helvik et al., 2006; John, Kreisman, & Pallett, 2012; Newman, Weinstein, Jacobson, & Hug, 1990). Therefore, activity limitations and participation restrictions in daily life for the affected individual should be taken into account as well (Ramkissoon & Cole, 2011). This implies a multidimensional approach (Solli & Da Silva 2012). Hence, in addition to pure-tone measures, other measures are often used to determine someone's hearing status. One such measure is the National Hearing Test (Smits, Kapteyn, & Houtgast, 2004), which measures the ability to understand speech in noise. This ability is poorly predicted by pure-tone thresholds. Besides the ability to understand speech in noise, self-reported activity limitations and participation restrictions are considered important to provide an ecologically valid assessment of an individual's hearing status. To assess perceived limitations and participation restrictions in real-life listening situations, patient-reported outcome measures (PROMs) can be used. PROMs are standardized, validated questionnaires that are administered to patients to assess their perceptions of their own functional status and well-being (Dawson, Doll, Fitzpatrick, Jenkinson, & Carr, 2010).

Several PROMs are available in the field of audiology. A well-known PROM is the Amsterdam Inventory for Auditory Disability and Handicap (AIADH; Kramer, Kapteyn, Festen, & Tobi, 1995). The items of the AIADH assess self-reported disability in everyday hearing. The AIADH

[a]Department of Otolaryngology–Head and Neck Surgery, Section Ear and Hearing, and EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands
[b]University of Amsterdam, the Netherlands
[c]Department of Epidemiology and Biostatistics, and EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands

Correspondence to Mariska Stam: mari.stam@vumc.nl

Editor: Nancy Tye-Murray
Associate Editor: Kathleen Cienkowski

Received April 30, 2015
Revision received August 26, 2015
Accepted September 10, 2015
DOI: 10.1044/2015_JSLHR-H-15-0156

Disclosure: The authors have declared that no competing interests existed at the time of publication.

has been widely used in different (clinical) populations. For example, Kramer et al. (1995), Neijenhuis, Stollman, Snik, & Van den Broek (2001), and Meijer, Wit, Tenvergert, Albers, and Kobold (2003) administered the AIADH to patients of an ear, nose, and throat department or an audiology clinic. Molander et al. (2013) used the AIADH in an internet-based hearing screening context in a sample of adults with and without hearing problems. Recently, dispensers of hearing aids in the Netherlands started using the AIADH as a diagnostic instrument for their clients. The AIADH has been embedded in a new protocol for hearing aid prescription and reimbursement. The original AIADH was developed in Dutch, but it has been translated into English (Kramer et al., 1995) and adapted into Spanish (Fuente, McPherson, Kramer, Hormazábal, & Hickson, 2012), Swedish (Hallberg, Hallberg, & Kramer, 2008), and Cantonese (Fuente, McPherson, Kwok, Chan, & Kramer, 2012). The AIADH is also included in a national ongoing internet-based longitudinal cohort study in the Netherlands, titled "Netherlands Longitudinal Study on Hearing" (NL-SH; e.g., Nachtegaal, Smits, et al., 2009; Stam et al., 2014).

Most of the currently used PROMs, including the AIADH, were developed using classical test theory (CTT), which has been the standard methodology for developing and analyzing psychometric properties of questionnaires over the past 70 years. CTT has well-known assumptions that are easy to meet in test data (Hambleton & Jones, 1993).

However, there are several disadvantages associated with CTT. An example is that under CTT, the actual pattern of responses to the items is discarded because the sum of these scores is the central outcome in this theory. Also, CTT is population dependent: Classical item and test statistics are only valid in the population in question; in populations with other score distributions, these statistics will be different (Iwata, 2014; Shultz, Whitney, & Zickar, 2014). The most important drawback associated with CTT is that it is assumed that measurement precision is identical for all ability levels, although for many PROMs, it seems likely that scores at the extremes are less reliable than scores in the middle of the range (Embretson & Reise, 2000).

A modern approach able to overcome these limitations is item response theory (IRT). IRT represents the relationship between an individual's item response and an underlying latent trait (theta, θ; Fraley, Waller, & Brennan, 2000). A latent trait is an unobservable entity that influences observable variables, such as item scores (Embretson & Reise, 2000). An example of an audiological latent trait is a person's perceived hearing ability. Whereas CTT is formulated only at the level of the total score, IRT provides an estimation of the latent trait, which takes into account the response pattern, and it provides separate parameters for both persons and items (Embretson & Reise, 2000). Also, unlike CTT, where reliability is typically represented by a single value (e.g., alpha, α), in IRT reliability is different at different levels of theta (Embretson & Hershberger, 1999). By using IRT, one is able to evaluate the psychometric properties of a scale and its items. When used appropriately, IRT produces precise and valid instruments with minimal response burden to the examinees (Edelen & Reeve, 2007) because it allows for the development of a computerized adaptive test (CAT). A CAT administers items to respondents via the computer, and the selection of each subsequent item for the individual is tailored to each individual's ability, so that the items are neither too difficult nor too easy (Embretson & Reise, 2000).

Despite the benefits of IRT, to our knowledge, only two audiological PROMs have been evaluated using IRT so far. The Communication Profile for the Hearing Impaired (Demorest & Erdman, 1987) was evaluated using IRT by Mokkink, Knol, van Nispen, and Kramer (2010) and by Demorest, Wark, and Erdman (2011). Mokkink et al. (2010) optimally shortened six Dutch scales of this PROM using IRT. Demorest et al. (2011) used IRT on this same PROM to select the best-performing items to be included in the development of a brief self-assessment screening instrument. Chenault, Berger, Kremer, and Anteunis (2013) also applied IRT. They used it for the Hearing Aid Rehabilitation Questionnaire (Hallam & Brooks, 1996) and demonstrated how application of IRT resulted in an efficient version of the Hearing Aid Rehabilitation Questionnaire for use in hearing screening and for the evaluation of interventions.

The AIADH has been proven to be a psychometrically sound instrument. However, because only CTT has been applied to the AIADH, no detailed psychometric information about individual item characteristics of the AIADH is available yet. The quality of this instrument and its potential for implementation in the clinic can be further improved by determining the item characteristics using IRT analyses. These characteristics are needed to be able to develop a CAT version in the future. A CAT version could potentially reduce response burden and improve the efficiency of this questionnaire. The aim of the present study was to investigate the properties of the AIADH using IRT. A large data set obtained through the NL-SH study, including both adults with hearing impairment and adults with normal hearing, was used for this study.

Method

Data Collection

Data were obtained from the ongoing web-based prospective cohort NL-SH study in the Netherlands (Nachtegaal, Kuik, et al., 2009; Nachtegaal, Smit, et al., 2009; Stam, Kostense, Festen, & Kramer, 2013; Stam et al., 2014). The NL-SH study uses a large convenience sample, including both adults with hearing impairment and adults with normal hearing. A wide range of variables is collected within the NL-SH, including the scores on the AIADH.

To enroll and inform participants, the NL-SH website (in Dutch: www.hooronderzoek.nl) was used. Data collection started in 2006. A variety of approaches was used to recruit participants. First, there is a link between the website of the National Hearing Test (for more information, see next section) and the website of the NL-SH. Anyone who is interested in testing his or her hearing ability and who performs the National Hearing Test is invited to participate in the NL-SH. In addition, a link to the study is permanently posted on other websites (e.g., the website of the hard-of-hearing patient organization). Also, flyers containing information about the study are regularly distributed in audiology clinics, at health fairs, and by dispensers of hearing aids in the Netherlands. In the past, advertisements were published in local newspapers.

Adults aged 18 to 70 years, from all over the Netherlands, are free to enroll themselves into the NL-SH study. The age range in the NL-SH study is restricted to 18–70 years because of the particular focus of this study on the effect of hearing impairment in young and middle-aged adults. After registration, participants receive a link to the online questionnaire, of which the AIADH is a part. Participants who do not respond within 1 week receive an e-mail reminder, and participants who then do not respond within 1 month receive a letter by regular mail. The NL-SH study is approved by the Medical Ethics Committee of the VU University Medical Center, Amsterdam, the Netherlands.

Measurements

The outcome measure in the present study is the AIADH (Kramer et al., 1995). The original AIADH is a 30-item PROM that assesses self-reported disability and handicap in everyday hearing. In this original questionnaire, if respondents had indicated they had difficulty in hearing (disability) in a particular situation, they were also asked how handicapped they felt. However, in the NL-SH study, only the "disability" was assessed, to limit response burden. According to the results of the initial CTT analyses on the AIADH impairment scores (Kramer et al., 1995), two questions were excluded (Items 18 and 30), resulting in a total of 28 items. These 28 items were administered in the NL-SH study. Together, they assess self-reported hearing ability, but they can be subdivided into five subscales, each representing a certain hearing domain: distinguishing sounds, auditory localization, intelligibility in noise, intelligibility in quiet, and detection of sounds. The response scale for each item is a 4-point Likert scale measuring how often the respondent is able to hear effectively in a specific situation: 0 = *almost always,* 1 = *frequently,* 2 = *occasionally,* or 3 = *almost never*. Higher scores denote higher impairment. Each item of the AIADH is accompanied by an image that visualizes the described hearing situation, thereby clarifying the circumstances of the specific listening situation (see Figure 1).

NL-SH participants were also instructed to perform the National Hearing Test (www.hoortest.nl). This test measures an individual's ability to understand speech in noise (Smits et al., 2004). Either headphones or speakers were allowed to be used, and participants were instructed to perform the test in a quiet room. In total, 23 digit triplets were presented against a background of stationary masking noise according to an adaptive (one-up, one-down) procedure. Listeners typed or clicked the digits on their keyboard or computer screen. The noise level was fixed, and the speech level varied. The speech reception threshold in noise (SRTn) was calculated by taking the average signal-to-noise ratio of the last 20 presentations, corresponding to 50% intelligibility. Based upon reference data used by Smits, Kramer, and Houtgast (2006), National Hearing Test scores were categorized into three categories of hearing: *good* (SRTn < -5.5 dB), *insufficient* ($-5.5 \leq$ SRTn ≤ -2.8 dB), and *poor* (SRTn > -2.8 dB). The aim of using these scores in the current study was to describe the study sample according to the National Hearing Test score categories. The sample consisted of adults with and without hearing impairment. Hence, the whole range of hearing abilities, from normal to poor hearing, was covered. The validity and reliability of the National Hearing Test have been proven to be adequate (Nachtegaal, Kuik, et al., 2009; Smits, Merkus, & Houtgast, 2006; Smits et al., 2004; Smits, Kramer, & Houtgast, 2006).

Statistical Analyses

Descriptive statistics were used to examine demographic characteristics, distributions of the National Hearing Test scores and AIADH scores, missing values, means, and standard deviations.

IRT: Checking its Assumptions

Before an IRT model may be applied, three core assumptions of the model need to be met: unidimensionality, local independence, and monotonicity. Unidimensionality means that the items in a questionnaire measure a single construct (Hays, Brown, Brown, Spritzer, & Crall, 2006). Therefore, all items of the AIADH should measure self-reported hearing ability. To test unidimensionality, a one-factor confirmatory factor analysis (CFA) was carried out on the polychoric correlation matrix of the AIADH items, which assess the degree of association between ordinal variables (Marcoulides & Raykov, 2006). The fit of the data to the one-factor CFA model was investigated by means of the following fit indices: Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and comparative fit index (CFI). Values indicating a good fit were RMSEA ≤ 0.06, and CFI and TLI ≥ 0.95 (Hétu et al., 1994). Furthermore, a principal components analysis (PCA), a form of exploratory factor analysis, on the polychoric correlations was carried out. In this way, we explored the amount of variance explained by the first factor, which should be at least 20% of the test variance for a questionnaire to be unidimensional. In addition, the difference in the magnitude of the proportion of explained variance between the first and the second factor was studied. A ratio greater than 4 is supportive for unidimensionality (Reeve et al., 2007).

Local independence implies that if the assumed abilities influencing the test performance are held constant, responses to any pair of items in the test are statistically independent (Swaminathan, Hambleton, & Rogers, 2006). In other words, if the self-reported hearing ability is held constant, responses to any pair of items of the AIADH should show no dependence. To examine local independence,

Figure 1. Example of one item of the Amsterdam Inventory for Auditory Disability and Handicap.

2. CAN YOU CARRY ON A CONVERSATION WITH SOMEONE IN A QUIET ROOM?

0 almost never 0 occasionally 0 frequently 0 almost always

the matrix of the residual correlations resulting from the one-factor CFA was examined. Item pairs with high residual correlations (>0.2) were considered as possibly locally dependent. In addition, local independence under the graded response model (GRM; described in detail in next paragraph) was studied using an IRT-based test of local independence: Yen's Q3 statistic (Yen, 1993). This statistic calculates residual item scores under the GRM (e.g., observed – expected response) and correlates these among items pairs (Embretson & Reise, 2000). Q3 values between 0.24 and 0.36 were considered as moderate deviations from model fit, and values of 0.37 and higher were considered as large deviations (Cohen, 2013).

Monotonicity means that the probability of endorsing an item response that reflects a higher ability increases as the underlying trait level increases (Reeve et al., 2007). So when adults have a higher hearing disability, they have a higher probability of endorsing the response option "almost never" able to hear or understand in the described situation. Monotonicity was studied using Mokken scaling, which is based on a nonparametric IRT model. In addition, the overall scalability coefficient H was used. With $0.30 \leq H < 0.40$ indicating a weak scale, $0.40 \leq H < 0.50$ indicating a medium scale, and $H \geq 0.50$ indicating a strong scale (Sijtsma, Debets, & Molenaar, 1990). Furthermore, scalability coefficients of each item were calculated. Coefficients needed to be higher than the lower bound of 0.3 to be considered as scalable.

IRT: The GRM

After checking the assumptions, an IRT model can be fit to the data. Various IRT models exist, but the one used in the present study was the GRM (Samejima, 1969), because it is applicable to items with ordered polytomous response categories (more than two categories), such as the four categories of the AIADH, ranging from "almost always" to "almost never." Furthermore, the GRM is easier to understand and illustrate than other models (Reeve et al., 2007). The GRM uses two types of item parameters to quantify the relationship between the latent trait and the item response: the discrimination parameter (a) and the threshold or difficulty parameter (b). The discrimination parameter (a) expresses the discriminative power of an item to mark differences between participants with similar scores on the latent trait. Each item has one or more difficulty parameters (b) to specify the location on theta where the examinee has a probability of at least 50% to select a given category (DeMars, 2010). The number of b parameters per item is equivalent to the number of response categories minus one (Furr & Bacharach, 2014), so the AIADH has three b parameters for each item.

To evaluate the fit of the GRM model, category response curves and item information curves were plotted. Category response curves represent the probability of a participant selecting a certain response category, given his or her level on the underlying trait (theta, θ; Reeve et al., 2007). Item information curves indicate the location on the latent trait scale at which an item is most reliable. The height of the curves (denoting more reliability) is a function of the a and b parameters, where higher curves indicate that someone's trait level can be measured more precisely. Item information curves indicate those items that are most useful for measuring different levels of the latent trait (Reeve et al., 2007).

Furthermore, marginal reliability was plotted. Marginal reliability is an index of reliability that is comparable to coefficient alpha, which is traditionally used as a measure of reliability in CTT (Scullard, 2007). It is the result of combining measurement error estimated at different points along the latent trait continuum into a single index. It corresponds to the average reliability of scores across all levels of theta (Ayearst & Bagby, 2011).

Software

Descriptive statistics were obtained using SPSS 20. CFA and IRT analyses were carried out using statistical package R version 3.1.0 (R Core Team, 2014). Different R libraries were used: CFA: Lavaan library (Rosseel, 2012), PCA: Psych library (Revelle, 2015), Mokken analysis: Mokken library (Van der Ark, 2007), and GRM: Ltm library (Rizopoulos, 2006).

Results

Demographics

The AIADH was completed and the National Hearing Test was carried out by 2,352 NL-SH participants between November 2006 and December 2013. The AIADH was completed fully by 99% of the respondents. In total, 26 respondents had missing items. These were excluded from Mokken and CFA analyses, because these do not allow missing values.

The age of the participants (36% men, 64% women) varied from 18 to 70 years ($M = 46.2$, $SD = 12.9$). In all, 23.6% of the participants reported using hearing aid(s), and 0.9% reported using a cochlear implant. Scores on the AIADH ranged from 0 to 3 (M = 0.8, SD = 0.7). National Hearing Test scores ranged from −13.3 to 4.0 dB signal-to-noise ratio (M = −4.6, SD = 3.6). According to the hearing categories defined by Smits, Kramer, and Houtgast (2006), 52.7% had good hearing ability, 21.7% had insufficient hearing ability, and 25.6% had poor hearing ability in noise.

Unidimensionality

Before the GRM model was applied to the data, the three core assumptions—unidimensionality, local independence, and monotonicity—were tested first. With regard to unidimensionality, the CFI and TLI fit indices showed a good fit to the one-factor CFA model. Both were greater than the required 0.95 (CFI: 0.99, TLI: 0.99). The RMSEA value of 0.09 was higher than the recommended value of 0.06. However, the PCA on the polychoric correlations showed that the first factor explained 67% of the variance, which is greater than the criterion of 20%. In addition, the second factor explained 5% of the variance. This resulted in a ratio of variance explained of the first to the second factor of about 13, which is higher than the required value of 4. These findings support the assumption that the items of the 28-item AIADH share a single common factor, and so the questionnaire can be considered unidimensional.

Local Independence

None of the 378 item pairs was possibly locally independent (i.e., multidimensional; criterion > 0.2). In addition, local independence under the GRM was studied using Yen's Q3. These results are shown in Table 1. Sixteen of the 378 item pairs (4.2%) had Q3 values that showed a moderate deviation from model fit (0.24 < Q3 < 0.36). Two item pairs had a large deviation from model fit (Q3 ≥ 0.37): Item 6 (recognize melodies in music or songs) and Item 24 (hear rhythm in music/songs), and Item 14 (understand news presenter on radio) and 20 (understand news presenter on television). Their Q3 values were 0.38 and 0.59, respectively. Due to the common content of these items, the local independence assumption might not hold strictly, but it might hold closely enough for IRT to be used advantageously in many practical situations (Kolen & Brennan, 2004). To investigate whether one of the items of these pairs might be considered redundant and might be removed, we followed the recommendations of Reeve et al. (2007). We fitted four new GRMs to the data. Each GRM left out one of the items 6, 24, 14, or 20. We then explored the changes in parameter estimates of the model after leaving out one of these four items. It appeared that the largest change in discrimination parameters (a) as well as difficulty parameters (b) was 0.10. Thus, the parameters of the model hardly changed due to removal of either Item 6 or Item 24, or Item 14 or Item 20. We therefore decided not to remove these items.

Monotonicity

Mokken scaling showed no significant violations in monotonicity. In addition, the H coefficient was 0.63, indicating a high scalability of the AIADH (H > 0.50 is considered as a strong scale). Furthermore, all item coefficients were higher than the required 0.3. Thus, the AIADH is able to show that when hearing ability decreases, the probability of endorsing the response category "almost never" will increase.

Table 1. Description of the 18 (out of 378) item pairs with possibly local independence under the graded response model using Yen's Q3 statistic.

Item	Short description	Item	Short description	Q3
1	Understand shop assistant in crowded shop	7	Carry on conversation with someone in crowded meeting	0.29
		25	Carry on conversation with someone in busy street	0.26
3	Hear from what direction car approaches	9	Hear from what direction a question is asked during meeting	0.29
		15	Look in right direction when called in the street	0.32
		27	Hear direction of a car horn	0.35
6	Recognize melodies in music or songs	24	Hear rhythm in music/songs	0.38
		29	Recognize and distinguish musical instruments	0.33
7	Carry on conversation with someone in crowded meeting	13	Carry on conversation with somebody in bus/car	0.30
		19	Follow conversation between few adults during dinner	0.29
		25	Carry on conversation with someone in busy street	0.34
9	Hear from what direction a question is asked during meeting	15	Look in right direction when called in the street	0.32
13	Carry on conversation with somebody in bus/car	19	Follow conversation between few adults during dinner	0.26
		25	Carry on conversation with someone in busy street	0.25
14	Understand news presenter on radio	20	Understand news presenter on television	0.59
15	Look in right direction when called in the street	27	Hear direction of a car horn	0.35
19	Follow conversation between few adults during dinner	25	Carry on conversation with someone in busy street	0.28
21	Hear from what corner when talked to in quiet house	27	Hear direction of a car horn	0.25
24	Hear rhythm in music/songs	29	Recognize and distinguish musical instruments	0.27

Note. 0.24 < Q3 < 0.36: moderate deviation from model fit; Q3 ≥ 0.37: large deviation from model fit.

Graded Response Model

The *a* and *b* parameters for each item of the AIADH are shown in Table 2. Items 9 (hear from what direction a question is asked during meeting) and 10 (hear somebody approach from behind) had the highest discrimination power, and Item 2 (carry on conversation in quiet room) had the lowest. The initial *b3* value of Item 2 was very high, indicating that the response category "almost never" was hardly chosen. Therefore, the categories "occasionally" and "almost never" were merged into one category. Still the *b2* parameter of Item 2 remained high (3.7). This indicates that a participant had to have a severe self-reported hearing disability before he or she would endorse the response category "occasionally" or "almost never" on the item about being able to carry on a conversation with someone in a quiet room (Item 2). Furthermore, the distribution of the category usage is shown in Table 2. Items 2, 16, and 23 displayed an underuse of category "almost never."

To visualize how the items fitted the GRM, category response curves for each item were plotted. See Figure 2 for examples of results of the category response curves. These curves show the probability of selecting one of the four response categories as a function of the hearing ability level. To explain, Figure 2 for Item 7 (carry on conversation with someone in crowded meeting) shows that an adult with a severe self-reported hearing disability (at the end of the right-hand side of the *x*-axis) has a high probability of endorsing the item with response category "almost never" (curve number 3) according to the AIADH total scale. An adult with moderate hearing disability has a higher probability of endorsing the response categories "occasionally" (curve number 2) or "frequently" (curve number 1). An adult with a relatively good self-reported hearing ability has a high probability of choosing "almost always" (curve number 0).

The category response curves for Item 7 (carry on conversation with someone in crowded meeting) are for a large part located on the left half of the hearing ability continuum (θ), more than, for example, the curves for Item 5 (recognize family members by voice). This indicates that not only those with self-reported hearing disability, but also those reporting no hearing difficulties are likely to endorse Item 7. The situation described in Item 7 is therefore considered a "difficult" listening situation. It addresses the

Table 2. Graded response model parameters of the Amsterdam Inventory for Auditory Disability and Handicap items and distribution of the response categories.

		Item parameters[a]				Distribution response categories[b] (%)			
Item	Short description	*a*	*b1*	*b2*	*b3*	0	1	2	3
1	Understand shop assistant in crowded shop	2.1	−0.3	0.8	2.2	41.4	27.2	24.7	6.7
2	Carry on conversation in quiet room	1.4	1.9	3.7[c]		86.1	12.0	1.9[c]	
3	Hear from what direction car approaches	2.5	−0.1	0.8	1.6	45.8	23.3	17.3	13.6
4	Hear cars passing by	2.5	0.5	1.6	2.8	63.2	23.5	11.3	2.0
5	Recognize family members by voice	2.0	0.8	2.0	3.1	68.5	22.5	7.4	1.6
6	Recognize melodies in music or songs	1.8	0.5	1.7	2.9	61.0	25.2	10.5	3.2
7	Carry on conversation with someone in crowded meeting	2.0	−1.1	0.3	1.8	21.6	35.1	32.4	10.9
8	Carry on telephone conversation in quiet room	2.0	1.3	2.4	3.2	80.6	14.6	3.3	1.4
9	Hear from what direction a question is asked during meeting	3.0	−0.4	0.5	1.3	37.3	25.8	19.7	17.2
10	Hear somebody approach from behind	3.0	−0.3	0.6	1.5	40.2	24.4	21.2	14.2
11	Recognize television presenter by voice	1.6	−0.5	0.9	2.2	36.9	33.2	20.9	9.0
12	Understand sung text	1.9	−0.7	0.6	1.7	30.9	33.1	23.0	13.1
13	Carry on conversation with somebody in bus/car	2.4	−0.7	0.4	1.7	30.0	30.6	28.5	10.8
14	Understand news presenter on radio	2.7	−0.1	0.8	1.5	44.8	25.0	17.0	13.3
15	Look in right direction when called in the street	2.5	−0.4	0.6	1.5	36.3	28.9	20.2	14.6
16	Hear household noises (running water, vacuuming)	1.9	0.9	2.1	3.7	70.4	21.7	7.2	0.7
17	Discriminate sound of car and bus	2.3	0.5	1.5	2.5	63.7	21.8	11.0	3.5
19	Follow conversation between few adults during dinner	2.1	−1.0	0.3	1.7	23.5	31.6	32.4	12.5
20	Understand news presenter on television	2.4	−0.1	0.8	1.7	45.3	24.9	18.3	11.5
21	Hear from what corner when talked to in quiet house	2.6	0.2	1.2	2.0	53.4	25.1	14.3	7.2
22	Hear doorbell at home	1.6	0.4	1.4	2.5	59.1	21.4	13.6	5.8
23	Distinguish between male and female voices	2.2	1.1	2.5	3.8	77.8	18.2	3.5	0.5
24	Hear rhythm in music/songs	1.8	0.9	2.1	3.3	71.6	19.7	6.9	1.8
25	Carry on conversation with someone in busy street	2.3	−0.9	0.4	1.8	24.8	34.2	31.3	9.7
26	Distinguish intonations in adults' voices	2.1	0.3	1.6	2.9	56.3	30.0	11.3	2.4
27	Hear direction of a car horn	2.4	−0.3	0.9	1.9	41.4	30.3	18.9	9.4
28	Hear birds sing outside	1.8	0.2	1.4	2.5	54.1	27.0	13.8	5.1
29	Recognize and distinguish musical instruments	1.7	−0.1	1.3	2.6	46.7	31.5	16.5	5.4

Note. In accordance with the results of Kramer et al. (1995), Items 18 and 30 were not assessed in the Netherlands Longitudinal Study on Hearing.

[a]*a* = discrimination parameter, *b* =difficulty parameters. [b]0 = *almost always,* 1 = *frequently,* 2 = *occasionally,* 3 = *almost never*. [c]The initial *b3* value of item 2 was very high; therefore, *b3* and *b2* for item 2 were merged into one category.

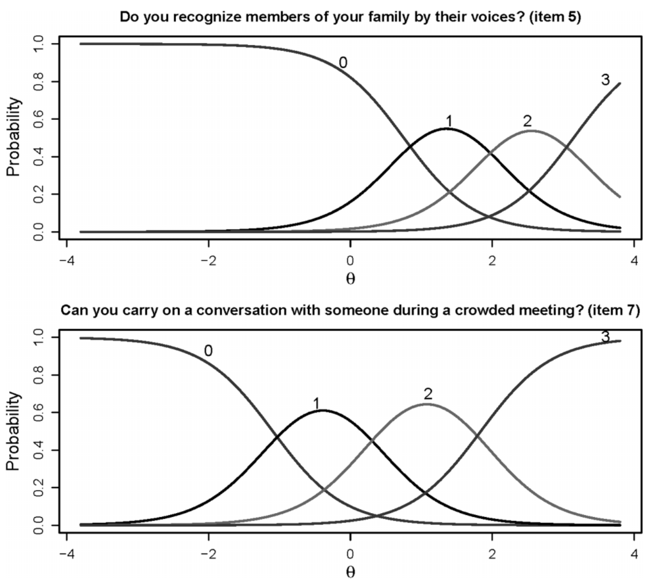


Figure 2. Category response curves for Items 5 (upper panel) and 7 (lower panel), displaying the probability of choosing one of the four response categories as a function of the self-reported hearing ability level (θ). High θ indicates severe self-reported hearing disability. 0 = *almost always,* 1 = *frequently,* 2 = *occasionally,* 3 = *almost never*.

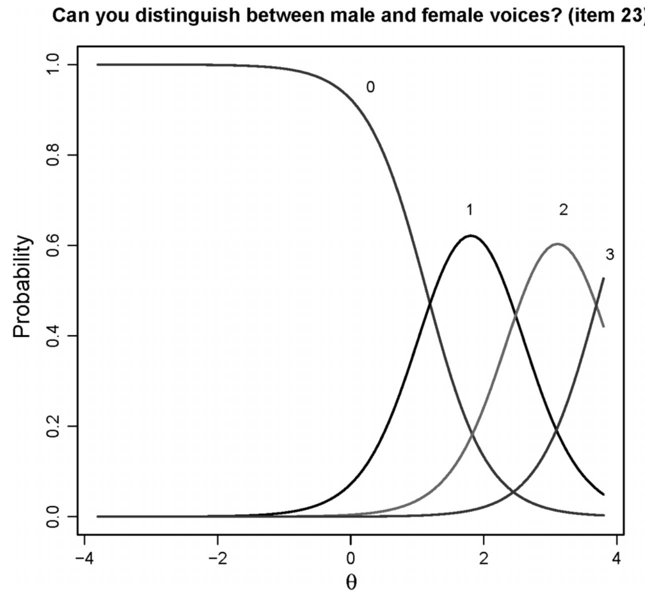


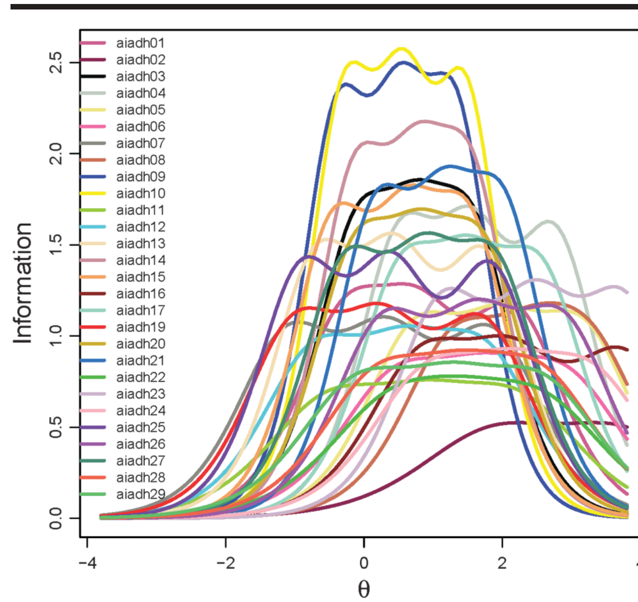
Figure 3. Right-centered category response curves of Item 23. High θ indicates severe self-reported hearing disability. 0 = *almost always,* 1 = *frequently,* 2 = *occasionally,* 3 = *almost never*.

ability to understand speech in noise, and the results suggest that understanding speech in noise can also be difficult for adults with normal hearing.

Almost all items had category response curves similar to those shown in Figure 2, where every response category has one region on the latent trait continuum where the probability of endorsing that category is highest. The intervals for Items 2 (understanding conversation in quiet room), 16 (hearing noises in household), and 23 (distinguish between male and female voices) were centered at the end of the right-hand side of the scale. This is in line with the results presented in Table 2, because these items have low frequencies for the category "almost never." The category response curves for Item 23 are presented in Figure 3 as an example. The response category "almost never" was only chosen by adults with self-reported severe hearing disability.

Item information curves were plotted to determine how informative each item was on the self-reported hearing ability continuum. The pattern in Figure 4 shows that, overall, the AIADH is most informative at the right-hand half of the scale (θ ≥ 0) because that is where the majority of the items had the highest information value. These items are mostly reliably measuring hearing disability in adults reporting hearing problems. The items with the highest information value at the right-hand end of the scale were Items 23 (distinguish between male and female voices) and 16 (hear household noises), with an information value of about 1.3 and 0.9, respectively. Overall, the AIADH items were less informative at the left-hand side of the scale, where adults with low theta values are located. These are people who do not experience hearing difficulties.

Figure 4. Item information curves of all Amsterdam Inventory for Auditory Disability and Handicap items as included in the Netherlands Longitudinal Study on Hearing . High θ indicates severe self-reported hearing disability. Higher information values indicate higher reliability; "aiadh01" to "aiadh29" in the graph legend refer to Items 1 to 29 of the Amsterdam Inventory for Auditory Disability and Handicap. In accordance with the results of Kramer et al. (1995), Items 18 and 30 were not assessed in the Netherlands Longitudinal Study on Hearing.



In Figure 5, we plotted the standard error as a function of the self-reported hearing ability. The lower panel shows the relative frequency of the latent disability variable. The upper panel shows the marginal reliability. The AIADH has a marginal reliability value of 0.8 for $\theta \geq -1.8$. In addition, the lower panel shows that the vast majority of the adults are located around these theta values of $\theta \geq -1.8$. These findings indicate that self-reported hearing ability is most reliably measured for adults who report having mild to moderate hearing disability. Reliability decreases (and standard error increases) at the lower values of θ (at the left-hand side), indicating that self-reported hearing ability is least reliably measured for adults who report having relatively good self-reported hearing ability.

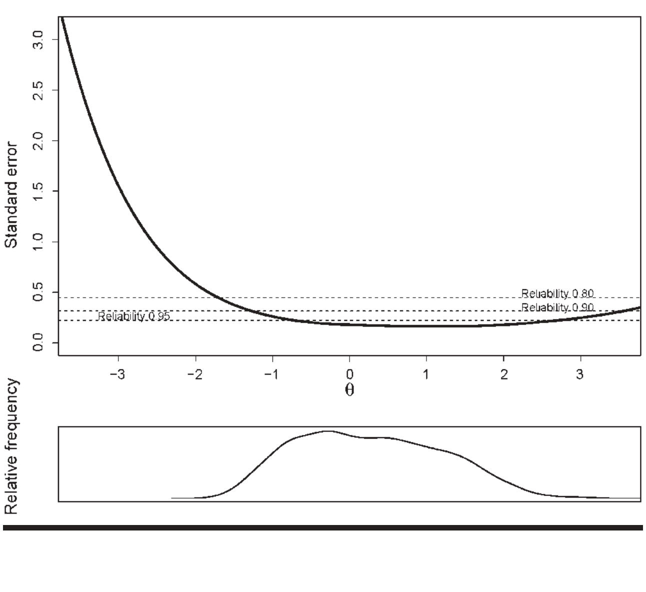
Discussion

The current study reevaluated the psychometric properties of the AIADH using a modern test theory, namely IRT, rather than CTT. It was shown that the items of the 28-item AIADH are scalable, which means that they can be hierarchically placed on the self-reported hearing ability continuum. Furthermore, the AIADH was least reliable for adults who reported having relatively good hearing ability, and it was most reliable for adults who reported having hearing disability. This is a clinically important finding, because the AIADH is being used by dispensers of hearing aids in the Netherlands as a clinical tool to assess the self-reported hearing disability of their clients.

Although the 28-item AIADH in general is more reliable for measuring self-reported hearing impairment, the item information curves showed that the highest curves are located around the middle of the hearing ability continuum. To assess severe self-reported hearing impairment, only a few moderately high item information curves are available in the AIADH (Item 23—distinguish between male and female voices, and Item 16—hear household noises). Consequently, small differences between adults with similar levels of self-reported hearing ability are less easily detected for respondents with severe self-reported hearing disability than for those with moderate self-reported hearing disability. It is questionable, however, whether more items that better target the whole continuum of self-reported severe hearing disability are required. Relatively fewer adults with hearing impairment will be located at the extremes of the continuum, because it is assumed that a population of adults with hearing impairment will follow a normal distribution; scores that are further away from the center have a lower frequency (Field, 2009). Although there may be more adults at the right end of the hearing ability continuum in a clinical study population than in the general population, it is doubtful whether the questionnaire should be made more sensitive to capture the specific problems related to severe hearing disability, just because of the earlier described argument. However, whenever the need to develop a questionnaire specifically aimed at capturing and distinguishing between severe and very severe levels of hearing disability becomes relevant, the results of the current study provide essential information, showing IRT parameters for each item individually. It is one of the few studies in audiology using IRT on PROMs.

Some discussion about the unidimensionality and local independence assumptions is needed. The 28-item AIADH was treated as a unidimensional scale in the present study, meaning that all items together measured self-reported hearing ability. However, during the development of the AIADH in 1995, a PCA with varimax and oblique rotation had shown that the questionnaire consisted of five factors (Kramer et al., 1995). To investigate whether these factors were also present in the NL-SH data set, we fitted a five-factor CFA model to the data. These analyses confirmed the five-factor model, and therefore the five factors will be retained. The underlying constructs are clinically relevant and important for the determination of individual hearing ability in a multidimensional manner. When the self-reported hearing ability is considered as a total score, some detailed information about activity limitations related to hearing could get lost. However, to be sure that, in spite of the five factors, it is legitimate to consider the AIADH as a unidimensional scale in the present study, we restudied the results of the PCA of the original 30-item AIADH (as published in Kramer et al., 1995). In that study, the first factor explained 39.8% of the variance, and the second factor explained 8%. This makes the ratio of the explained variance of the first to the second factor greater than the required ratio of 4 for unidimensionality. This supports the use of the original 30-item AIADH as a single construct, equal to what we observed in the current study. Therefore, the AIADH can be considered unidimensional, and its five factors can be retained to clinically assess the self-reported hearing ability of the patient in different hearing domains.

Figure 5. The upper curve represents the standard error as a function of θ. Horizontal lines represent marginal reliability. The lower curve represents the relative frequency estimates of the distribution of θ. Again, high θ indicates severe self-reported hearing disability.



With regard to the local independence assumption, two item pairs (Items 6 and 24, and Items 14 and 20) were possibly locally dependent under the GRM model. However, this is not unexpected, because it is commonly understood that item responses are rarely strictly unidimensional. In many cases, multidimensionality is due to the heterogeneous item content that is required to properly represent the complexity of health constructs. Acknowledging this fact, researchers have focused on methods of exploring whether data are "unidimensional enough" for IRT application (Reise, Morizot, & Hays, 2007). To check whether the possible local dependency effected the model, we followed the item-in–item-out approach recommended by Reeve et al. (2007). The results showed that the GRM model was not influenced by the possible local dependency and therefore the AIADH can be considered "unidimensional enough" for IRT application.

Psychosocial consequences of hearing impairment, such as decreased quality of life, cannot be predicted from audiometric data alone (Hallberg et al., 2008). The use of a subjective rating scale additional to an "objective" measure may be necessary and the only way to increase the likelihood of truly capturing the phenomenon of interest (Kayes & McPherson, 2010). This emphasizes the need for the use of PROMs in audiology. The current results confirm that self-reported hearing disability can be measured with high measurement precision. This further emphasizes the importance of IRT models, as describing item functioning along a continuum cannot easily be achieved with CTT (Reeve & Fayers, 2005).

Strengths and Limitations

The strengths of the current study are the large sample of 2,352 participants and the inclusion of both adults with hearing impairment and adults with normal hearing in the sample, which provides more information than a purely clinical sample. Furthermore, it is one of the few studies on self-reported measures in audiology using IRT instead of CTT. A limitation of the current study is that the participants were rather young. Inclusion of older respondents in this study might have resulted in a different set of item characteristics. Future IRT research including older respondents is therefore recommended.

Implications

The present study is an important step toward the actual implementation and further development of PROMs in audiology, because IRT analyses provide a promising basis for the development of a CAT. In a CAT, the computer selects items on the basis of the answers to previous items. Instead of giving each examinee the same fixed questionnaire, CAT item selection adapts to the ability level of the individual examinee. An optimal item is the item with the highest informative value for the item discrimination parameter (van der Linden & Glas, 2000). In other words, each adult will only receive the items that are most reliable for their degree of ability. By administering only the items that are relevant to a given individual, the ability can be determined accurately and efficiently, resulting in a major reduction in the time required to administer or complete instruments (Gibbons et al., 2008).

CAT is not that developed yet in audiology, but in other disciplines, CAT is common, and its usage has been found promising. For example, in cardiology, Abberger et al. (2013) implemented a CAT to measure self-reported anxiety in cardiovascular patients. Implementation of this CAT into clinical routine showed that a short processing time was needed to measure anxiety with high measurement precision. The results of the current study are a promising basis for the development of a CAT version of the AIADH: an efficient (screening) tool with low response burden and accurately assessed self-reported hearing disability, in both clinical and experimental audiology settings. It would be interesting to carry out a simulation study with real patient data to evaluate the algorithm of the CAT, and in addition, to perform an experimental study to evaluate and validate a CAT version in clinical routine.

Conclusions

The IRT analyses showed accurate description of item functioning of the AIADH along the self-reported hearing ability continuum. Different reliability values were determined for any degree of self-reported hearing ability, instead of just one general reliability value for the whole scale. All 28 AIADH items contributed reliably to the measurement of self-reported hearing impairment in adults. None of the items had to be removed according to our analyses. These findings provide the basis for the development of a CAT version of the AIADH: an efficient (screening) tool with advantages for dispensers of hearing aids, audiology clinics, and ear, nose, and throat departments, as well as their clients or patients, because of a quicker, more precise, and personalized procedure with which to assess self-perceived hearing impairment.

Acknowledgments

The authors thank the participants of the Netherlands Longitudinal Study on Hearing. The Netherlands Longitudinal Study on Hearing was financially supported by the Heinsius Houbolt Foundation, and partly funded by Phonak AG, Switzerland. The current study is financially supported by Health Insurance Netherlands.

References

Abberger, B., Haschke, A., Wirtz, M., Kroehne, U., Bengel, J., & Baumeister, H. (2013). Development and evaluation of a computer adaptive test to assess anxiety in cardiovascular rehabilitation patients. *Archives of Physical Medicine and Rehabilitation, 94,* 2433–2439.

Ayearst, L. E., & Bagby, R. M. (2011). Evaluating the psychometric properties of psychological measures. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of Assessment and Treatment*

Planning for Psychological Disorders (2nd ed., pp. 23–61). New York, NY: Guilford.
Chenault, M., Berger, M., Kremer, B., & Anteunis, L. (2013). Quantification of experienced hearing problems with item response theory. *American Journal of Audiology, 22*(1), 252–262.
Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. London: Academic Press.
Dawson, J., Doll, H., Fitzpatrick, R., Jenkinson, C., & Carr, A. J. (2010). The routine use of patient reported outcome measures in healthcare settings. *British Medical Journal, 340,* 464–467.
DeMars, C. (2010). *Item response theory*. Oxford, United Kingdom: Oxford University Press.
Demorest, M. E., & Erdman, S. A. (1987). Development of the communication profile for the hearing impaired. *Journal of Speech and Hearing Disorders, 52,* 129–143.
Demorest, M. E., Wark, D. J., & Erdman, S. A. (2011). Development of the screening test for hearing problems. *American Journal of Audiology, 20*(2), 100–110.
Diao, M., Sun, J., Jiang, T., Tian, F., Jia, Z., Liu, Y., & Chen, D. (2014). Comparison between self-reported hearing and measured hearing thresholds of the elderly in China. *Ear and Hearing, 35,* e228–e232.
Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16*(Suppl. 1), 5–18.
Embretson, S. E., & Hershberger, S. L. (1999). *The new rules of measurement: What every psychologist and educator should know*. New York, NY: Psychology Press.
Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
Field, A. (2009). *Discovering statistics using SPSS*. London: Sage Publications.
Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology, 78,* 350.
Fuente, A., McPherson, B., Kramer, S. E., Hormazábal, X., & Hickson, L. (2012). Adaptation of the Amsterdam Inventory for Auditory Disability and Handicap into Spanish. *Disability and Rehabilitation, 34,* 2076–2084.
Fuente, A., McPherson, B., Kwok, E. T. T., Chan, K., & Kramer, S. E. (2012). Adaptation of the Amsterdam Inventory for Auditory Disability and Handicap into Cantonese. *The Australian and New Zealand Journal of Audiology, 32,* 115–126.
Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: An introduction* (2nd ed.). Los Angeles, CA: Sage Publications.
Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., . . . Immekus, J. C. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services, 59,* 361–368.
Granberg, S., Pronk, M., Swanepoel, D. W., Kramer, S. E., Hagsten, H., Hjaldahl, J., . . . Danermark, B. (2014). The ICF core sets for hearing loss project: Functioning and disability from the patient perspective. *International Journal of Audiology, 53,* 777–786.
Hallam, R. S., & Brooks, D. N. (1996). Development of the Hearing Attitudes in Rehabilitation Questionnaire (HARQ). *British Journal of Audiology, 30,* 199–213.
Hallberg, L. R. M., Hallberg, U., & Kramer, S. E. (2008). Self-reported hearing difficulties, communication strategies and psychological general well-being (quality of life) in patients with acquired hearing impairment. *Disability and Rehabilitation, 30,* 203–212.
Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Instructional Topics in Educational Measurement, 12*(3), 38–47.
Hays, R. D., Brown, J., Brown, L. U., Spritzer, K. L., & Crall, J. J. (2006). Classical test theory and item response theory analyses of multi-item scales assessing parents' perceptions of their children's dental care. *Medical Care, 44*(11), S60–S68.
Helvik, A. S., Jacobsen, G., Wennberg, S., Arnesen, H., Ringdahl, A., & Hallberg, L. R. M. (2006). Activity limitation and participation restriction in adults seeking hearing aid fitting and rehabilitation. *Disability and Rehabilitation, 28,* 281–288.
Hétu, R., Getty, L., Philibert, L., Desilets, F., Noble, W., & Stephens, D. (1994). Development of a clinical tool for the measurement of the severity of hearing disabilities and handicaps. *Journal of Speech-Language Pathology and Audiology [French], 18*(2), 83–95.
Iwata, N. (2014). Cultural distinctiveness in response bias. In M. F. Dollard, A. Shimazu, R. Bin Nordin, P. Brough, & M. R. Tuckey (Eds.), *Psychosocial factors at work in the Asia Pacific*. Dordrecht, the Netherlands: Springer.
John, A. B., Kreisman, B. M., & Pallett, S. (2012). Validity of hearing impairment calculation methods for prediction of self-reported hearing handicap. *Noise and Health, 14,* 13–20.
Kayes, N. M., & McPherson, K. M. (2010). Measuring what matters: Does "objectivity" mean good science? *Disability & Rehabilitation, 32,* 1011–1019.
Kolen, M., & Brennan, R. L. (2004). *Test equating, linking, and scaling: Methods and practices*. New York, NY: Springer.
Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Tobi, H. (1995). Factors in subjective hearing disability. *International Journal of Audiology, 34,* 311–320.
Marcoulides, G. A., & Raykov, T. (2006). *A first course in structural equation modeling*. Mahwah, NJ: Erlbaum.
Meijer, A. G., Wit, H. P., Tenvergert, E. M., Albers, F. W., & Kobold, J. P. M. (2003). Reliability and validity of the (modified) Amsterdam Inventory for Auditory Disability and Handicap. *International Journal of Audiology, 42,* 220–226.
Mokkink, L. B., Knol, D. L., van Nispen, R. M., & Kramer, S. E. (2010). Improving the quality and applicability of the Dutch scales of the Communication Profile for the Hearing Impaired using item response theory. *Journal of Speech, Language, and Hearing Research, 53,* 556–571.
Molander, P., Nordqvist, P., Öberg, M., Lunner, T., Lyxell, B., & Andersson, G. (2013). Internet-based hearing screening using speech-in-noise: Validation and comparisons of self-reported hearing problems, quality of life and phonological representation. *BMJ Open, 3,* e003223.
Nachtegaal, J., Kuik, D. J., Anema, J. R., Goverts, S. T., Festen, J. M., & Kramer, S. E. (2009). Hearing status, need for recovery after work, and psychosocial work characteristics: Results from an internet-based national survey on hearing. *International Journal of Audiology, 48,* 684–691.
Nachtegaal, J., Smit, J. H., Smits, C. A. S., Bezemer, P. D., van Beek, J. H., Festen, J. M., & Kramer, S. E. (2009). The association between hearing status and psychosocial health before the age of 70 years: Results from an internet-based national survey on hearing. *Ear and Hearing, 30,* 302–312.
Neijenhuis, K. A., Stollman, M. H., Snik, A. F., & Van den Broek, P. (2001). Development of a Central Auditory Test Battery for adults. *International Journal of Audiology, 40,* 69–77.
Newman, C. W., Weinstein, B. E., Jacobson, G. P., & Hug, G. A. (1990). The Hearing Handicap Inventory for Adults: Psychometric adequacy and audiometric correlates. *Ear and Hearing, 11,* 430–433.

Pronk, M., Deeg, D. J., Smits, C., van Tilburg, T. G., Kuik, D. J., Festen, J. M., & Kramer, S. E. (2011). Prospective effects of hearing status on loneliness and depression in older persons: Identification of subgroups. *International Journal of Audiology, 50,* 887–896.

R Core Team. (2014). *R: A language and environment for statistical computing* (version 3.1. 0). Vienna, Austria: R Foundation for Statistical Computing [Computer software].

Ramkissoon, I., & Cole, M. (2011). Self-reported hearing difficulty versus audiometric screening in younger and older smokers and nonsmokers. *Journal of Clinical Medicine Research, 3,* 183–190.

Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In P. Fayers & R. Hays (Eds.), *Assessing quality of life in clinical trials: Methods of practice*. New York, NY: Oxford University Press.

Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., . . . Cella, A. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care, 45*(5 Suppl. 1), S22–S31.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*(Suppl. 1), 19–31.

Revelle, W. (2015). *psych: Procedures for psychological, psychometric, and personality research* (R package version 1.5.1) [Computer software]. See also http://www.personality-project.org/r/psych

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software, 17,* 1–25.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48,* 1–36.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* [Monograph]. Richmond, VA: Psychometric Society.

Scullard, M. G. (2007). *Application of item response theory based computerized adaptive testing to the strong interest inventory*. Ann Arbor, MI: ProQuest.

Shultz, K. S., Whitney, D. J., & Zickar, M. J. (2014). *Measurement theory in action: Case studies and exercises, second edition*. New York, NY: Routledge.

Sijtsma, K., Debets, P., & Molenaar, I. W. (1990). Mokken scale analysis for polychotomous items: Theory, a computer program and an empirical application. *Quality and Quantity, 24,* 173–188.

Smits, C., Kapteyn, T. S., & Houtgast, T. (2004). Development and validation of an automatic speech-in-noise screening test by telephone. *International Journal of Audiology, 43,* 15–28.

Smits, C., Kramer, S. E., & Houtgast, T. (2006). Speech reception thresholds in noise and self-reported hearing disability in a general adult population. *Ear and Hearing, 27,* 538–549.

Smits, C., Merkus, P., & Houtgast, T. (2006). How we do it: The Dutch functional hearing-screening tests by telephone and internet. *Clinical Otolaryngology, 31,* 436–440.

Solli, H. M., & Da Silva, A. B. (2012). The holistic claims of the biopsychosocial conception of WHO's international classification of functioning, disability, and health (ICF): A conceptual analysis on the basis of a pluralistic–holistic ontology and multidimensional view of the human being. *Journal of Medicine and Philosophy, 37,* 277–294.

Stam, M., Kostense, P. J., Festen, J. M., & Kramer, S. E. (2013). The relationship between hearing status and the participation in different categories of work: Demographics. *Work: A Journal of Prevention, Assessment and Rehabilitation, 46,* 207–219.

Stam, M., Kostense, P. J., Lemke, U., Merkus, P., Smit, J. H., Festen, J. M., & Kramer, S. E. (2014). Comorbidity in adults with hearing difficulties: Which chronic medical conditions are related to hearing impairment? *International Journal of Audiology, 53,* 392–401.

Swaminathan, H., Hambleton, R. K., & Rogers, H. J. (2006). 21 Assessing the fit of item response theory models. *Handbook of Statistics, 26,* 683–718.

Van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software, 20,* 1–19.

van der Linden, W. J., & Glas, C. A. (2000). *Computerized adaptive testing: Theory and practice*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30,* 187–213.