# Assessing Readiness for Hearing Rehabilitation

# Assessing Readiness for

# Hearing Rehabilitation

**DISSERTATION**

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. dr. L.L.G. Soete,
in accordance with the decision of the Board of Deans,
to be defended in public on Thursday 22 Octobre 2015, at 16:00 hours

by

Michelene N. Chenault

UNIVERSITAIRE
PERS MAASTRICHT

**Supervisors**

      Prof. dr. Martijn P.F. Berger
      Prof. dr. Bernd Kremer


**Co-Supervisor**

      Dr. Lucien J.C. Anteunis


**Assessment committee**

      Prof. dr. Gerard J.P. van Breukelen (Chair)
      Dr. Erwin L.J. George
      Prof. dr. Sophia E. Kramer, VU Medical Center, Amsterdam
      Prof. dr. Robert J. Stokroos
      Prof. dr. Henrica C.W. de Vet, VU Medical Center, Amsterdam

# Contents

## Abbreviations

| | |
|---|---|
| CTT | Classical Test Theory |
| DIF | Differential Item Functioning |
| DTF | Differential Test Functioning |
| EFA | Exploratory Factor Analysis |
| GRM | Graded Response Model |
| ICC | Item Characteristic Curve |
| IRT | Item Response Theory |
| IRTLR | Item Response Theory Likelihood-ratio |
| LR | Logistic Regression |
| OLR | Ordinal Logistic Regression |
| PRO | Patient Reported Outcome |
| PTA | Pure Tone Audiometry |
| PTABE | Pure Tone Audiometry Best Ear |
| VAS | Visual Analogue Scale |
| WHO | World Health Organization |

# Chapter 1

General introduction

# Introduction

Among persons aged 55 or more, 30% were found to have a hearing loss of 30 dB or more in their better ear.[1] Only half of these hearing impaired persons had ever consulted their general practitioner about their hearing and only half of these consulters had become hearing aid users. Davis et al.[2] concluded that hearing aid uptake to remedy age-related hearing loss occurs on average 10 years later than when it would be optimal and that systematic screening among persons aged 55 to 74 years would not only provide substantial benefit but would be cost effective as well. Hearing aid uptake cannot be equated with successful hearing rehabilitation since hearing aids are often left in a drawer unused.[3] A number of studies have discussed factors which influence whether or not a person with hearing impairment seeks help, whether or not help-seeking results in hearing aid uptake and what factors influence satisfied hearing aid use. A great portion of this research has been described in two relatively recent literature reviews.

The review by Knudsen et al.[4] (2010) examined literature regarding help-seeking behaviour for hearing loss, hearing aid uptake, hearing aid use and hearing aid satisfaction. These four factors are linked to the path towards successful hearing rehabilitation. Each of these factors is more strongly associated with self-reported hearing difficulties than with audiometric outcomes, providing evidence that self-report of hearing problems should be at the basis of the assessment for hearing aid candidacy. While the studies reviewed did not contradict each other regarding their findings, a clear consensus is lacking as to how relevant outcomes should be defined and/or operationalized, pointing out the need to identify critical outcome domains. A number of potentially influential factors such as the role of self-efficacy, social support and coping were discussed with the suggestion that more qualitative research be conducted on interpersonal processes.

With the objective of providing direction for future research, the literature study by Pronk et al.[5] (2011) focused on interventions subsequent to hearing screening. One of the topics addressed was the potential role of screening interventions to increase the hearing impaired person's knowledge and awareness of his/her hearing loss. The notion is that screening could be implemented as an intervention to motivate the hearing impaired person to take action to remedy hearing difficulties. Milstein and Weinstein[6] (2002) applied the Health Belief Model and the Trans-theoretical Stages-of-Change Model when implementing a screening with follow-up, with the rationale being that contributing to the individual's knowledge of the consequences of hearing loss will bring about a change in behavior.

The present dissertation, besides contributing to the ongoing discussion regarding barriers to rehabilitation, is directed towards the development of a screening tool to determine an individual's need and readiness for hearing rehabilitation. Within this context the concept of the patient journey and the methodologies related to Patient

Reported Outcomes (PROs) and Item Response Theory (IRT) are applied. These topics are introduced in the following along with a short description of the questionnaire used in Chapters 2 through 6.

## Patient journey

Some studies have adopted a qualitative approach to making sense of the process the individual undergoes from becoming hearing impaired to becoming a satisfied hearing aid user. Manchaiah et al.[7] (2011) described the patient journey towards rehabilitation as having seven stages: 1) pre-awareness; 2) awareness; 3) movement; 4) diagnostics; 5) rehabilitation; 6) self-evaluation; and 7) resolution. Since a satisfied hearing aid user is one who has arrived at the last stage or resolution, other hearing impaired persons would be located at any one of the 6 preceding stages.

## Patient Reported Outcomes (PROs)

Clearly the challenge remains of identifying factors and how they interrelate to influence behaviour on this patient journey. Since these factors are either subjective experiences or attitudes and cannot be measured directly they are referred to as latent factors or traits. If these latent factors could be identified and quantified it might be possible to assign clearer specifications of the passage along the patient journey. Questionnaires are tools to quantify such latent factors. When questionnaires are applied to health related matters they are known as Patient Reported Outcomes (PROs). The application of PROs has been growing over the past several of decades. Exploratory Factor Analysis (EFA) has been the standard method for deriving scales composed of questionnaire items to quantify latent variables.

## Hearing Aid Rehabilitation Questionnaire (HARQ)

The HARQ was originally developed by Hallam and Brooks[8] (1996) as a tool to aid in the rehabilitation process. The original HARQ consists of 40 questions with three response alternatives: 'agree'= 1; 'partly agree'= 2 and 'disagree'= 3 distributed over seven scales, three of which pertain to attitudes regarding hearing and four to attitudes regarding hearing aids. The hearing scales were labeled: 1) personal distress/inadequacy, 2) hearing loss stigma and 3) minimization of hearing impairment while the hearing aid scales were labeled: 1) hearing aid stigma, 2) aid not wanted, 3) pressure to be assessed and 4) positive expectation of aid. The responses to items within each scale were summed with a higher scale score indicating a greater manifestation of that dimension. For the present dissertation the HARQ questionnaire, administered to a sample representative of a screening population, is re-evaluated for use as an assessment instrument for hearing problems and hearing aid attitudes in screening and clinical settings.

## Item Response Theory (IRT)

In the past decade there has been a surge in the application of IRT to further refine and calibrate scales obtained through exploratory factor analysis. There are two general approaches, namely Classical Test Theory (CTT) and IRT, to measure latent traits. CTT is the traditional approach whereby the numerically coded responses to items within a questionnaire scale are summed to obtain a score quantifying the underlying trait. CTT disregards the fact that responses to individual items within a scale differ relative to the trait being assessed. In contrast, IRT models the probability of a response to each item as a function of the underlying trait, so that the contribution of each item relative to various trait levels can be examined. Hayes and Lipscomb[9] (2007) discussed the advantages of IRT methodology, being that IRT models provide a better depiction of actual response patterns. Moreover, IRT estimates are a more accurate representation of the latent trait being estimated and provide the possibility of gaining sensitivity in cross-sectional comparisons and when assessing health changes over time. An additional advantage is that IRT methodologies have been developed to detect bias in questionnaire responses.

The response scale of the questionnaire items being considered determines the types of IRT models which can be estimated. Since the HARQ questionnaire items are on a three point ordinal scale, the Graded Response Model (GRM) proposed by Samejima[10,11] (1969, 1997) is appropriate. For each item within a scale there are two types of parameters to be estimated: discrimination and thresholds. These parameters are estimated relative to a range of values representing the magnitude of the subjective experience or attitude, known as the latent trait. An item which has a higher discrimination parameter can discriminate more sharply between response categories. Higher estimated threshold parameters for a particular item indicate that higher trait values are required for transition from one response category into the adjacent higher category. Given that the items considered here have three response categories, two thresholds are estimated per item. These two thresholds correspond to the transition from "disagree" to "partly agree" and from "partly agree" to "agree" or to the transition from "agree" to "partly agree" and from "partly agree" to "disagree", depending on how the responses are linked to the underlying trait.

## Summary of Chapters 2 through Chapter 6

In Chapter 2 the results of administering the HARQ is presented. Exploratory factor analysis is applied separately to the hearing and hearing aid items using the polychoric correlations among the items responses. Reliability analysis is performed to determine the consistency of each scale. The derived scale scores are examined relative to hearing aid use and hearing impairment defined according to hearing aid reimbursement criteria at the time of data collection.

In Chapter 3 the assumptions required for GRM, namely monotonicity, uni-dimensionality and local independence, are evaluated for the two hearing scales derived in Chapter I. GRM parameter estimations for each of the items in each scale for the three point response scale and for the two point response scale, resulting from merging the categories "partly agree" and "agree", are compared and discussed.

Chapter 4 is devoted to investigating bias in the two hearing scales relative to hearing impairment and hearing aid use whereby methodology from the Differential Test Framework is applied using the procedure suggested by Teresi and Fleishman[12] (2007). This procedure includes the IRTLR (Item Response Theory Likelihood Ratio) method developed by Thissen et al.[13] (1993), and the OLR (Ordinal Logistic Regression) which is an extension of the LR (Logistic Regression) method developed by Zumbo[14] (1999). Biased items are removed from their respective scales.

In Chapter 5, the GRM models of the three scales which emerged from the HARQ hearing aid items are estimated. Model fit at the item level is assessed[15,16]. The presence of bias is investigated following the methodology adhered to in Chapter 4.

In Chapter 6, hearing aid use and the WHO classifications of hearing impairment are subjected to discriminant analysis. In this analysis the five scales, reduced as a result of the investigation of bias in Chapters III and IV, are considered to distinguish between hearing aid users and non- hearing aid users, between persons who are severely and moderately hearing impaired, moderately and mild hearing impaired and mild and non-hearing impaired. Because of heterogeneity in group variances, a non-linear discriminant function is estimated. A stepwise procedure is adopted to determine which combinations of scales and items together provided the optimal classification of the groups being compared.

In Chapter 7, the main results are discussed with a summary of the findings of Chapters 2 through 6. Recommendations for future research are proposed. A valorization for this research study is given.

# References

1. Duijvestijn JA, Anteunis LJC, Hoek CJ, Van den Brink RHS, Chenault MN, Manni JJ. Help-seeking behavior of hearing-impaired persons aged ≥55 years; effect of complaints, significant others and hearing aid image. *Acta Otolaryngol* 2003;123:846-850.

2. Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I. Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models. *Health Technol Assess* 2007;11:1-294.

3. Joore MA, Van Der Stel H, Peters HJ, Boas GM, Anteunis LJ. The Cost effectiveness of hearing-aid fitting in the Netherlands. *Arch Otolaryngol Head Neck Surg* 129:297-304.

4. Knudsen LV, Oberg M, Nielsen C, Naylor G, Kramer SE. Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids: a review of the literature. *Trends Amplif* 2010;14:127-54.

5. Pronk M, Kramer SE, Davis AC, Stephens D, Smith PA, Thodi C, Anteunis LJ, Parazzini M, Grandori F. Interventions following hearing screening in adults: a systematic descriptive review. *Int J Audiol* 2011; 50:594-609.

6. Milstein D, Weinstein, BE. Effects of information sharing on follow-up after hearing screening for older adults. *JARA* 2002;35:43-58.

7. Manchaiah VK, Stephens D, Meredith R. The patient journey of adults with hearing impairment: the patients' views. *Clin Otolaryngol* 2011;36:227-234.

8. Hallam RS, Brooks DN. Development of the Hearing Attitudes in Rehabilitation Questionnaire (HARQ). *Br J Audiol* 1996;30:199-213.

9. Hays RD, Lipscomb J. Next steps for use of item response theory in the assessment of health outcomes. *Qual Life Res* 2007;6:195-199.

10. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 1969;No. 17.

11. Samejima F. Graded response model. In W.J. van der Linden & R.K. Hambleton (Eds.) Handbook of modern item response theory. Springer. 1997:85-100.

12. Teresi JA, Fleishman JA. Differential item functioning and health assessment. *Qual Life Res* 2007;16: 33-42.

13. Thissen D, Steinberg L, Wainer H. Detection of differential item function using the parameters of item response models. In P.W. Holland & H. Wainer (Eds.) Differential item functioning. Hillsdale NJ: Lawrence Erlbaum, Inc. 1993:67-113.

14. Zumbo BD. A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic regression modeling as a Unitary Framework for Binary and Likert-type (Ordinal) item Scores. Ottawa. On: Directorate of Human Resources Research and Evaluation, Department of National Defense. 1999.

15. Orlando M, Thissen D. Further investigation of the performance of S-$\chi^2$: An item fit index for use with dichotomous item response theory models. *APM* 2013;27:289-298.

16. Cai L, Thissen D, du Toit S. IRTPRO 2.1 for Windows, SSI, Inc. 2012.

# Chapter 2

Evaluation of the Hearing Aid Rehabilitation
Questionnaire in Dutch: examination of its
psychometric properties and potential
use as a screening instrument

MN Chenault
LJC Anteunis
MPF Berger

2

## Abstract

Items pertaining to hearing and hearing aids from the Hearing Aid Rehabilitation Questionnaire were applied to a heterogeneous sample of Dutch patients aged 55 years and more to evaluate their potential use in hearing screening. Subjects aged 55+ were recruited from a large general practitioners practice to participate. Three groups were formed: a group of 63 persons with a hearing aid, a group of 64 without a hearing aid but with sufficient hearing impairment to qualify for hearing aid reimbursement, and a group of 85 non-hearing impaired persons. Factor and reliability analyses revealed a structure with two scales regarding hearing, namely functionality and social hearing and three scales pertaining to hearing aids, namely hearing aid stigma, pressure to be assessed and not wanting a hearing aid. Scale validity was assessed with pure tone averages over the frequencies 1, 2 and 4 kHz and with a visual analogue scale for subjective hearing. The derived scales can be applied reliably in audiological assessment in an adult hearing screen setting to detect experienced hearing problems as well as attitudes related to hearing and hearing aids.

## Introduction

It has been shown that hearing impairment has a profound effect upon an individual's quality of life.[1] Even though hearing aid use improves quality of life[2] and hearing aid fitting is a cost effective health intervention,[3] there is a great disparity between those who might benefit and those who actually pursue hearing rehabilitation. This disparity has been attributed to a number of factors which are either related to behavioural strategies and/or the presence of stigma's and attitudes towards hearing impairment and hearing aids.[4,5] Moreover it has been shown that cognitive ability is related to successful hearing aid use.[6]

Loss of hearing ability, specifically presbyacusis, is an accompaniment of the aging process and thus is usually gradual whereby the individual goes through stages in which limitations in hearing are experienced and responded to with various behaviours including coping, avoidance, denial and help-seeking. Various studies have attempted to describe the process from the onset of hearing impairment to either successful or unsuccessful hearing rehabilitation. Manchaiah et al.[7] performed a qualitative study to map the patient journey from the onset of hearing impairment to successful hearing rehabilitation and identified seven stages: i) pre-awareness, ii) awareness, iii) movement, iv) diagnostics, v) rehabilitation, vi) self-evaluation and vii) resolution, whereby the first two stages are the most relevant for hearing screening. The pre-awareness stage is characterized by the individual not recognizing his/her hearing limitations while others do recognize it. The individual may experience frustration and exhibit avoidance behaviour. In the awareness stage the person has recognized some hearing limitations and attempts to identify situations in which problems are most prominent. The stages thereafter pertain to help-seeking behaviour, clinical assessment and initiation of rehabilitation.

Hearing screening facilitates the help-seeking behaviour as it confronts the individual with the facts of his/her hearing impairment.[8] Stephens and Kramer[9] pointed out that audiological assessment in screening should be based on the person's experienced difficulties in hearing rather than on audiometric outcomes. Based on an extensive literature study, including a number of studies reporting the necessity of enhancing audiological assessment with patient reporting, Knudsen et al.[10] concluded that self-reported auditory difficulty is possibly more important than clinical audiometric measurements for hearing assessment. Other studies have investigated an approach towards screening which combines audiometric testing and patient reporting. Yueh et al.[11] conducted a scientific review of screening methods for adult hearing loss and concluded that using a tone-emitting otoscope conjointly with the Hearing Handicap Inventory for the Elderly-Screening Version (HHIE-S)[12] was the best strategy. They pointed out that audiometry assesses physiological hearing loss while the HHIE-S addresses social and emotional handicap.

2

Pronk et al.[13] conducted an extensive and comprehensive literature study on hearing screening. In particular they discussed interventions, which follow failing a hearing screen, which usually is limited to referral for hearing aid fitting. However there are studies, which have implemented interventions including education and training regarding hearing impairment and how to cope with it. Hickson et al.[14] found that communication programs should be integrated in audiological rehabilitation and supplement traditional interventions such as hearing aid fitting.

Health screening programs are meant to filter out those individuals who would benefit from intervention. Without screening, hearing rehabilitation depends solely on help-seeking behaviour with the hearing impaired person having already transgressed into the movement stage. Assessment of experienced limitations and difficulties should therefore accompany audiometric testing in hearing screening. Patient answered questionnaire items facilitate the quantification of limitations experienced as a result of hearing impairment and thus the burden experienced. Methodology regarding the application of questionnaire items for assessing health outcomes has been developing at a fast pace with increasing emphasis being placed on health functional aspects rather than clinical measurement.

The validity and reliability of a questionnaire designed to measure health outcomes must be examined like any other measurement tool. Cronbach and Meehl[15] were pioneers in the development of classical test theory regarding the reliability and validity of tests and questionnaires. Since then, others such as Terwee et al.[16] have suggested additional criteria to be met for questionnaires targeting health status. Hyde[17] specifically investigated standards for subjective rating scales used in audiological rehabilitation and made suggestions regarding various aspects of validity and reliability. In particular, he emphasized the necessity of examining existing measures with regard to i) purpose: whether the objective is assessment of hearing ability or ability to successfully undergo rehabilitation, ii) context: whether the items are appropriately formulated for the population of interest and for the hearing-related trait(s) being measured and iii) responsiveness: whether the items are formulated to minimize unusable responses such as *non-applicable* while obtaining responses which truly reflect the traits being measured.

Screening will filter out persons already recognizing their disability but also those who are in the pre-awareness stage. Screening should be supplemented with questionnaire items which quantify experienced burden. In this way not only those recognizing their disability but also those who may be beginning to experience limitations can be discerned. Screening should also assess attitudes, which potentially thwart help-seeking behaviour. Items are then needed which are directed towards measuring how the individual views others with hearing disability and/or hearing aids. Screening items addressing such topics may also serve as an intervention mediating, that is confronting the individual with heretofore unacknowledged hearing difficulties, or moderating how

2

the individual views his/her own hearing impairment. Such items are included in the Hearing Aid Rehabilitation Questionnaire (HARQ).[18]

The HARQ was developed and validated in the UK to assess attitudes to serve as a tool in the rehabilitation process of middle-aged to elderly persons, aged 60 years or more. It is not clear how widespread the HARQ has been applied in hearing rehabilitation, although studies can be mentioned which used parts of the HARQ to measure initial attitudes before hearing intervention. Hickson et al.[14] used the 20 hearing impairment items as one scale to measure the impact of initial attitudes in the Active Communication Education program for elderly persons, while Jerram and Purdy[19] used the hearing aid stigma scale consisting of 9 items to assess pre-fitting attitudes. Furthermore, the domains determined by the HARQ have been adopted in the development of other questionnaires such as the questionnaire developed by Meister et al.[20] addressing pre-fitting expectations. The HARQ addresses a wider range of topics than the HHIE-S, including attitudes, perceived benefit, experienced social limitations and social pressure. The HHIE-S consists of 10 questions addressing the emotional and social impact of hearing difficulty. The 40-item HARQ addresses not only these aspects but also coping and whether the individual has been pressured to be assessed, and issues related to hearing aids such as stigma and whether hearing aids are considered useful and desirable. Two other questionnaires which have been developed by Cox and Alexander[21,22] for use in hearing rehabilitation deserve mention. The Satisfaction with Amplification with Daily Life (SADL)[21] measures hearing aid satisfaction while its sister questionnaire Expected Consequences to Hearing Aid Ownership (ECHO)[22] has the same items as the SADL but the wording is in the future tense. Although these instruments are more comprehensive in how they measure (expected) satisfaction the authors found the items in the HARQ addressing hearing aid expectation appropriate for the present study.

In the present paper the results of the administration of a translated version of the HARQ in a Dutch sample will be investigated to determine whether it (or parts of it) can be applied in adult hearing screening to detect experienced hearing problems and assess attitudes related to hearing and hearing aids. In contrast to these earlier studies, which used the HARQ for persons at the start of the rehabilitation process, our focus is on screening and therefore calls for a more heterogeneous sample. Hearing aid users have also been included to form a base for comparison with the non-hearing aid users. This latter group is composed of persons with varying degrees of hearing impairment, e.g. persons who would and would not qualify for hearing aid reimbursement according to their level of hearing impairment and thus resembling a screening population. The validation is therefore twofold with attention for both the translation and its application in a screening setting. Since the HARQ was originally designed to assess persons at the beginning of the rehabilitation process, not only different response patterns are expected but also missing values due to lack of suitability for some in the target audience. Furthermore, the HARQ will be examined relative to the patient

**2**

journey towards hearing rehabilitation. Although hearing aid users would not typically be screened for hearing, they are being included in this study for comparison purposes as these persons are at the end of the patient journey towards hearing rehabilitation.

## Materials and methods

### Original development of the Hearing Aid Rehabilitation Questionnaire

The HARQ items are on a three-point response scale: "disagree"; "partly agree" and "agree" with "non-applicable" or "don't know" as additional response categories. Factor analysis was originally performed by Hallam and Brooks[18] separately for the hearing impairment items and the hearing aid items although these items are presented in mixed form in one questionnaire. They derived three scales addressing hearing loss: i) personal distress/inadequacy, ii) hearing loss stigma and iii) minimization of hearing impairment and four scales addressing hearing aids: i) hearing aid stigma, ii) aid not wanted, iii) pressure to be assessed and iv) positive expectation of aid. The items were summed so that the higher the sub-scale score, the more manifest the dimension, with four items recoded, which had reversed. The Manual of the HARQ was revised in 2008 and is available on request from the author R. Hallam (polpresa@gmail.com).

For the present study, the original HARQ was translated into Dutch by a Dutch native speaker and then independently back translated into English by an English native speaker to verify retention of the original meaning. The original and the back-translated English versions were then checked by both translators, first independently and then together, leading to minor changes in formulations before it was found suitable for use in this study.

### Study group

Our heterogeneous sample consisted of persons recruited from the patient population of the Health Centre Neerbeek, a large general practitioner (GP) group practice in the province of Limburg in the Netherlands with 1302 registered patients aged 55 and more. This group was stratified by age to obtain random samples of relatively equal size for three age categories: 55-64, 65-74 and 75+ years, so that the sample would include individuals who are either employed, are retired but relatively active and those who are clearly elderly, since this might influence how important hearing is in their daily life. A letter was sent by their GP to a random sample of 557 patients requesting them to participate in a hearing study, followed by telephone contact two weeks later resulting in 271 participants who were scheduled for an audiogram, from which participants were included in one of three study groups. The aim was to obtain three approximately equal study groups: individuals with moderate hearing impairment and using a hearing

aid: the aided group, individuals with a moderate hearing impairment not owning a hearing aid: the unaided group, and individuals without hearing impairment and without a hearing aid: the normal hearing group. Prior to 2013 Dutch health insurance granted (partial) reimbursement for hearing aids at a best ear pure tone average (BEPTA) of ≥ 35 dB. Therefore this criterion was applied to define our groups in the recruitment stage. This cut-off point of 35 dB over the frequencies 1, 2 and 4 kHz does not coincide with hearing impairment classifications adhered to by the European Commission (EC) and the World Health Commission (WHO). The EC (1996) and the WHO (2001) classify hearing impairment according to average thresholds for the frequencies 0.5, 1, 2 and 4 kHz in the better ear. The EC classification is no impairment for less than 20 dB, mild for 21-39, and moderate for 40-69, while according to the WHO classification: 25 dB or less is considered no impairment, 26-40 dB as slight or mild, 41-60 dB as moderate, 61-80 as severe and more than 80 dB as profound. Persons with mild to moderate hearing loss were therefore considered the target hearing impaired participants for the present study. As only 14 moderately hearing-impaired hearing aid users could be recruited from the GP clinic, this group was supplemented with 46 hearing aid users from the records of the Audiological Department of the Maastricht University Hospital. These 46 hearing aid users were recruited in the same way via an invitation letter from their own GP followed up by a phone call. The sample resulting consisted of 212 persons, 106 males and 106 females with three study groups: the *normal* group (n=85) with hearing impairment less than 35 dB in the better ear for the frequencies 1, 2, 4 kHz, the *unaided* group (n=64) with hearing important of 35 dB or more in the better ear without a hearing aid and the *aided* group (n=63) with a hearing aid. The distribution according to the WHO hearing impairment classification within these groups is presented in Figure 2.1.
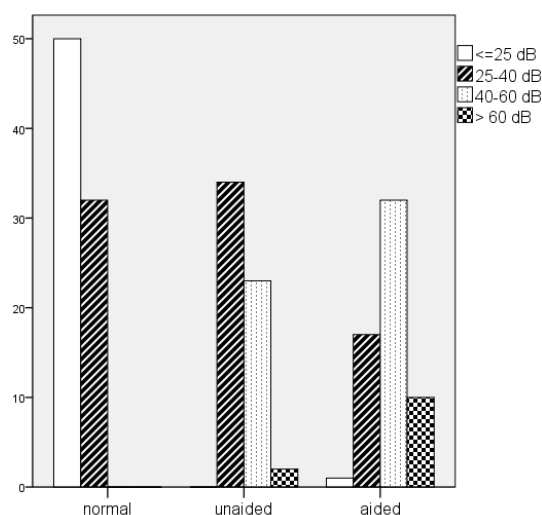


Figure 2.1    World Health Organization hearing impairment classification per hearing group in counts.

2

## Questionnaire administration

The Dutch HARQ was administered by one of two research assistants in a personal interview setting where the participant completed the questionnaire in written form and the research assistant was present to clarify any questions the participant may have. A visual analogue scale (VAS) was used to register the participant's subjective hearing (dis)ability, without a hearing aid for the hearing aid users, a score of 0 being the worse and 100 being the best possible hearing ability. This scale was presented as a horizontal ruler of 10 cm with 10 marks from 0 to 100 and the participant was asked to provide a mark to indicate his/her subjective hearing. Audiometric measurement was scheduled prior to the interview. The present study was approved by the University Hospital Medical Ethics Committee and informed consent was obtained from the participants.

## Statistical methodology

The 21 HARQ hearing impairment and 20 HARQ hearing aid items were analyzed separately just as in the original study.[18] As the items are on a 3-point ordinal scale, polychoric correlation matrices were obtained separately for the hearing and hearing aid items, on the basis of which factor analysis was performed. The calculation of polychoric correlations is a technique for estimating the correlation between two theorized normally distributed continuous latent variables, from two observed ordinal variables.[23]

Statistics were obtained to determine whether the assumptions regarding sampling adequacy were met. Values for Kaiser-Meyer Olkin Measure of Sampling Adequacy (KMO) <0.5 are unacceptable while values above 0.7 are preferable. Bartlett's Test of Sphericity tests the null hypothesis that there are no correlations between the items so that a highly significant result indicates that there are correlations between items. Scree plots and parallel analysis were applied to determine the number of factors. The *Scree test*[24] is a graph of the eigenvalues plotted against the factors. The number of factors is indicated by the point above the elbow: the cut-off point between major and trivial factors. Parallel analysis[25] is based on comparing eigenvalues obtained from the data and those obtained from generating random values. Thus a comparison is made between what is observed and what would be expected if there were no factor structure present. As the obtained scales were expected to correlate, oblique (oblimin) rotations were performed to identify factor structures and these were substantiated with orthogonal rotation (varimax). A loading of 0.4 or more was considered as a significant contribution of the item to the factor.[26] The identified factor structures were compared to those found by Hallam and Brooks.[18] Cronbach's of each scale was calculated. Items were discarded if they compromised the reliability coefficient.

Criterion or concurrent validity of hearing scales were investigated by comparing scale outcome values with BEPTA 1, 2, 4 kHz and the VAS assessing subjective rating of

hearing ability, for the hearing aid group this was answered without a hearing aid. Criterion group validity was assessed by comparing scale scores of the three hearing groups and according to the WHO hearing impairment classification.

Since the items are on an ordinal scale the software program PRELIS was used to obtain polychoric correlation matrices to be used in the factor analysis. Data analysis was conducted with SPSS 18.

## Results

### Sample characteristics

Age and gender characteristics of the whole sample and separately for the three groups are presented in Table 2.1 along with pure tone averages (PTA), asymmetry, and the original HARQ scale scores. Mean age did not differ significantly between the groups, however there were relatively more males in the aided group (Chi-square=8.63, 2 df, $p$=0.013). One way ANOVA confirmed that PTA's differed significantly between the three groups, all three pair-wise comparisons being statistically significant ($p$<0.0005) after Bonferroni adjustment, with as expected the aided group having the greatest mean hearing impairment and the normal hearing group the lowest. There were relatively more persons with asymmetry exceeding 10 dB for the frequencies 0.5, 1, 2, 4 kHz in the hearing aid group (chi-square=8.30, 2 df, $p$=0.016). Likewise, the VAS hearing score was highest for the normal hearing group and lowest for the aided group, all three pair-wise comparisons being statistically significant ($p$<0.0005) after Bonferroni adjustment. Sub-scale scores obtained for the original HARQ scale conformed to ranges given by Hallam and Brooks.[18] Moreover the scale scores obtained per hearing (aid) group were plausible with the aided group having the highest mean values for the scales personal distress and pressure to be assessed and the lowest mean values for the other four scales. In contrast the scale scores for hearing aid stigma and positive expectation of aid were highest for the impaired but unaided group.

### Polychoric correlations

In the calculation of the polychoric correlation coefficients of the 21 hearing items, PRELIS gave a warning regarding the accuracy of the estimated correlation coefficients for an item from the original minimization of hearing impairment scale: "Hearing is not a serious problem for me". This item was therefore discarded leaving 20 hearing items for the factor analysis. In calculating the polychoric correlations of the hearing aid items, PRELIS gave a warning regarding the accuracy of coefficient estimations of two of the three items from the original positive expectation scale. These two items pertained to the time that would be required to adjust to using a hearing aid. The third item from this scale, which addressed expected benefit, had a high percentage of missing values.

2

It was therefore decided to exclude these three items. Estimation of correlations with the seventh item of the hearing aid stigma scale: "Many people don't know how to react to you when you have a hearing aid" also presented problems and therefore this item was discarded as well leaving a total of 16 hearing aid items for factor analysis.

Table 2.1 Percentage of males and asymmetry >10 dB, mean (SD), *median (range)* of age, pure tone averages, visual analogue scale-hearing, original Hearing Aid Rehabilitation Questionnaire scale scores for hearing groups: i) normal: non-hearing impaired; ii) unaided: hearing impaired without hearing aid (best ear pure tone average 1, 2, 4 kHz >35 dB); iii) aided: hearing aid.

| | total sample (n=212) | normal (n=85) | unaided (n=64) | aided (n=63) |
|---|---|---|---|---|
| % males | 50.0 | 41.2 | 46.9 | 65.1 |
| age in years | 71.1 (8.6) | 69.1 (8.8) | 74.8 (7.4) | 70.0 (8.8) |
| | *71 (59,96)* | *68 (56,96)* | *76 (58,92)* | *69 (56,92)* |
| BEPTA 1,2,4 kHz | 38.3 (15.9) | 23.5 (6.6) | 44.7 (8.8) | 52.3 (13.3) |
| | *37 (7,88)* | *25 (7,33)* | *43 (35,68)* | *51 (25,88)* |
| BEPTA 0.5, 1,2,4 kHz | 36.3 (14.1) | 23.6 (6.0) | 41.7 (8.2) | 48.3 (12.6) |
| | *35 (11,78)* | *24 (11,39)* | *40 (29,61)* | *49 (24,78)* |
| WEPTA 0.5,1,2,4 kHz | 43.7 (17.0) | 29.8 (10.6) | 47.9 (12.0) | 58.4 (13.5) |
| | *41 (13,94)* | *28 (13,68)* | *44 (31,83)* | *60 (35,94)* |
| Asymmetry dB (0.5,1,2,4  kHz) | 7.4 (8.5) | 6.2 (7.9) | 6.3 (7.4) | 10.1 (9.8) |
| | *5 (0,41)* | *4 (0,41)* | *4 (0,35)* | *6 (0,39)* |
| % Asymmetry > 10 dB | 19.4 | 13.4 | 15.3 | 31.7 |
| VAS hearing score | 65.7 (20.5) | 79.4 (14.5) | 61.8 (17.3) | 51.8 (19.3) |
| | *70 (5,100)* | *80 (45,100)* | *60 (15,100)* | *50 (5,90)* |
| Personal distress | 16.3 (5.5) | 13.8 (3.5) | 16.4 (5.7) | 19.9 (5.6) |
| | *14 (11,33)* | *13 (11,28)* | *14 (11,33)* | *19 (11,33)* |
| Hearing loss stigma | 6.0 (1.7) | 6.2 (1.6) | 5.7 (1.2) | 6.1 (2.1) |
| | *5 (5,15)* | *5 (5,10)* | *5 (5,10)* | *5 (5,15)* |
| Minimization of hearing impairment | 12.4 (2.5) | 13.0 (1.9) | 12.8 (2.8) | 11.0 (2.5) |
| | *12 (6,18)* | *12 (9,18)* | *13 (6,18)* | *11 (6,16)* |
| Hearing aid stigma | 12.6 (3.0) | 12.8 (2.7) | 12.7 (2.8) | 12.2 (3.4) |
| | *12 (9,25)* | *12 (9,20)* | *12 (9,21)* | *11 (9,25)* |
| Aid not wanted | 8.5 (2.5) | 9.4 (1.8) | 9.3 (2.4) | 6.2 (1.7) |
| | *9 (5,15)* | *9 (5,15)* | *9 (5,15)* | *5 (5,11)* |
| Pressure to be assessed | 3.6 (1.2) | 3.2 (0.7) | 3.5 (1.1) | 4.3 (1.6) |
| | *3 (3,9)* | *3 (3,7)* | *3 (3,7)* | *4 (3,9)* |
| Positive expectation aid | 7.0 (1.8) | 7.0 (1.8) | 7.2 (1.6) | 6.8 (2.1) |
| | *7 (3,9)* | *7 (3,9)* | *7 (4,9)* | *7 (3,9)* |

BEPTA: best and WEPTA worse ear pure tone average. Asymmetry: absolute difference PTA 0.5,1,2,4 kHz left/right ears.

## Reliability analysis of the Hearing Aid Rehabilitation Questionnaire hearing items

Assumptions required for factor analysis were met regarding sampling adequacy (KMO=0.8) and Bartlett's test of sphericity was highly significant (*p*<0.0005). The scree plot suggested a two-factor solution with the elbow more clearly between the second

and third eigenvalues and this was supported by parallel analysis. The item: "I've come to regard whatever hearing difficulties I may have as a problem not worth bothering about" had a low loading on both factors. Of the remaining 19 items there were 11 items with loadings >0.4 on the first factor and 10 on the second factor, thus with 2 items loading on both. After examination of the contents of the two factors it was decided that the 11-item-factor could be labelled *functionality* while the 10-item-factor could be labelled *social hearing*. The *functionality* pertains to experienced difficulties in functioning while *social hearing* pertains to difficulties experienced in a social context. Factor loadings of the items of the derived hearing scales are presented in Table 2.2 (for each item the loading obtained by Hallam and Brooks is presented in parentheses). Scale sum scores were obtained by adding the response values of 1, 2, and 3 for disagree, partly agree and agree whereby the items with negative loadings were recoded (1=3, 2=2 and 3=1). A Cronbach's alpha of 0.85 was obtained for the 11 item *functionality* scale and 0.84 for the *social hearing* scale.

Table 2.2    Derived scales & factor loadings of HARQ hearing items.

| Items according to Hallam & Brooks scales and loadings ( ) | functionality | social hearing |
|---|---|---|
| **Personal Distress** | | |
| 1  It sometimes depresses me when I cannot follow a conversation (.57) | .85 | |
| 2  I dread meeting new people since becoming hearing impaired  (.59) | .47 | .59 |
| 3  I find myself avoiding company because conversation is too much effort (.71) | .55 | .58 |
| 4  In a conversational group I keep quiet for fear of saying the wrong thing (.67) | .66 | |
| 5  When several people are chatting, it bothers me that I often lose the thread of the conversation (.67) | .75 | |
| 6  It really upsets me when I realise I've got the "wrong end of the stick" in conversation (.70) | .61 | |
| 7  By and large I am able to hear without difficulty (-.51) | -.86 | |
| 8  I can hear well enough when I really concentrate (-.50) | -.70 | |
| 9  My poor hearing sometimes makes me feel really inadequate (.71) | | .76 |
| 10 My hearing loss makes me feel isolated from other people (.74) | | .74 |
| 11 I have to admit that deep down I feel restricted by my hearing loss (.67) | | .63 |
| **Minimization of Hearing Impairment** | | |
| 1  I think I've overcome any hearing difficulties I might have through my own efforts (.49) | .50 | |
| 2  Difficulty in hearing is not of major concern to me at the moment (.46) | .69 | |
| 3  My hearing problems are really quite minor (.41) | .70 | |
| 4  By and large I am able to hear without difficulty (.44) | | |
| 5  I've come to regard whatever hearing difficulties I may have as a problem not worth bothering about (.76) | | |
| 6* Hearing is not a serious problem for me (.52) | | |
| **Hearing Loss Stigma** | | |
| 1  When you have hearing difficulties, other people ignore you (.55) | | .67 |
| 2  As I see it, I am less of a person because of my hearing difficulty (.62) | | .65 |
| 3  I am sure that some people think I am stupid (.73) | | .71 |
| 4  I get the feeling that other people find it a strain to talk to me (.54) | | .70 |
| 5  Some people avoid me because of my hearing ability (.72) | | .75 |
| *Total explained variance:  57.9%* | *36.1* | *30.8* |
| *Cronbach's alpha* | *.85* | *.84* |

*not included

## Reliability analysis of Hearing Aid Rehabilitation Questionnaire hearing aid items

The assumptions required for factor analysis were met: KMO=0.60 and Bartlett's test of sphericity was highly significant ($p$<0.0005). The scree plot and parallel analysis supported a three-factor solution for the hearing aid items. There were two items, which loaded on two factors and one, which loaded on all three. There was one item, "I think the behind-the-ear aids are really quite small and inconspicuous" which had a marginal (< 0.4) loading on the social pressure scale. The obtained factor structures for the hearing aid items could be labelled *aid stigma*, *social pressure*, and *aid unwanted* (Table 2.3).

Cronbach's 's of 0.62, 0.61, and 0.49 were obtained respectively for the 6 item *aid stigma*, the 7 item *social pressure* and the 5 item *aid unwanted* scales.

Table 2.3    Derived scales and factor loadings of HARQ hearing aid items.

| Items according to Hallam & Brooks scales and loadings ( ) | aid stigma | social pressure | aid unwanted |
|---|---|---|---|
| **Hearing Aid Stigma** | | | |
| 1  I think the behind-the-ear aids are really quite small and inconspicuous  (-.53) | | | |
| 2  If I wear aid,  people will probably think I'm a bit stupid (.63) | .70 | .40 | |
| 3  It would embarrass me to have to wear a hearing aid (.61) | .40 | .53 | .54 |
| 4  I think people react differently to you when you are wearing a hearing aid (.60) | .68 | | |
| 5  I would stand out in a crowd wearing a hearing aid (.67) | .79 | | |
| 6  I think that if you wear a hearing aid people tend to ignore you (.72) | .68 | | |
| 7* Many people don't know how to react to you when you have a hearing aid (.66) | | | |
| 8  It would make me feel old to wear a hearing aid (.65) | .80 | | |
| 9  From what I know, hearing aids don't help a great deal (.52) | | | .77 |
| **Aid Not Wanted** | | | |
| 1  I don't really want a hearing aid (.74) | | | .71 |
| 2  I think that wearing a hearing aid would help me when meeting strangers (.56) | | | .72 |
| 3.  I am willing to try a hearing aid but don't think an aid will be of much help to me (-.71) | | | -.58 |
| 4  I don't consider it important to be assessed for a hearing aid (.72) | | -.59 | |
| 5  My hearing is not so bad that I need a hearing aid (.72) | | -.73 | |
| **Pressure to be assessed** | | | |
| 1  I feel I have been pressured into having my hearing assessed (.85) | | .71 | |
| 2  I have come about my hearing in order to please someone else (.79) | | .72 | |
| 3  It is due to pressure from my family or friends that I am having my hearing assessed (.86) | | .78 | |
| *Total explained variance : 56.4%* | *20.6* | *25.7* | *10.2* |
| *Cronbach's alpha* | *.62* | *.61* | *.49* |

*not included

## Validity analysis

For each of the scales a higher scale score indicates a greater problem. Inspection of the mean values of the five derived scales (Table 2.4) shows that the *functionality*, *social hearing* and the *social pressure* scales were lowest for the normal hearing group and highest for the aided group, while the *aid stigma* scale was lowest for the aided group and highest for the normal hearing group. All pair-wise comparisons for the *functionality* scale relative to either the hearing group or the WHO classification were statistically significant even after Bonferroni correction ($p<0.0005$). For the *social hearing* scale the greatest difference was observed between the normal and aided group ($p=0.031$ but after Bonferroni correction this was 0.092). The *social hearing* scale means differed significantly for all pair-wise comparisons of the WHO classifications with the exception of the no impairment and slight/mild categories. The scale *aid stigma* was highest for the normal group and lowest for the aided group but these differences were small while no trend could be observed relative to the WHO classification. The *social pressure* scale scores were lowest for the normal and highest for the aided group and this trend was also observed for the WHO classification. All pair-wise comparisons were statistically significant except between the normal and unaided group ($p=0.052$). The scale *aid unwanted* was highest for the unaided and lowest for the aided group. Pair-wise comparisons showed statistically significant differences between the normal hearing and unaided group ($p=0.05$) and between the normal hearing and aided groups ($p<0.0005$), however after Bonferroni correction the difference between the normal and unaided groups was no longer significant. The *aid unwanted* scores decreased relative to the level of impairment as defined by the WHO classification but these differences were not statistically significant.

Table 2.4    Mean (sd), *median (range)* of derived scale sum scores, according to sample hearing group and WHO hearing classification.

| | Functionality | Social hearing | Aid stigma | Social pressure | Aid unwanted |
|---|---|---|---|---|---|
| Total Sample | 17.1 (5.7) | 12.8 (3.9) | 7.8 (2.3) | 10.2 (2.4) | 7.2 (2.0) |
| | *15 (11,33)* | *11 (10,29)* | *7 (6,16)* | *9 (7,19)* | *7 (5,13)* |
| Sample group classification | | | | | |
| Normal | 13.8 (3.1) | 12.2 (3.2) | 8.0 (2.2) | 8.9 (1.2) | 7.4 (1.6) |
| | *13 (11,26)* | *11 (10,23)* | *7 (6,14)* | *9 (7,13)* | *7 (5,12)* |
| Unaided | 17.1 (5.6) | 12.7 (3.7) | 7.8 (2.1) | 9.8 (2.0) | 8.0 (2.2) |
| | *16 (11,32)* | *12 (10,24)* | *7 (6,14)* | *9 (7,15)* | *8 (5,13)* |
| Aided | 21.6 (5.4) | 13.7 (4.7) | 7.6 (2.5) | 12.5 (2.2) | 6.2 (1.8) |
| | *22 (11,33)* | *12 (10,29)* | *6 (6,16)* | *12 (8,19)* | *5 (5,12)* |
| WHO classification | | | | | |
| ≤ 25 dB | 13.9 (3.2) | 12.4 (3.2) | 8.2 (2.4) | 8.9 (1.4) | 7.4 (1.6) |
| | *13 (11,26)* | *11 (10,21)* | *7 (6,14)* | *9 (7,15)* | *7 (5,12)* |
| 26-40 dB | 15.7 (4.5) | 11.7 (2.4) | 7.5 (1.8) | 9.9 (2.3) | 7.2 (1.8) |
| | *14 (11,27)* | *11 (10,21)* | *7 (6,13)* | *9 (7,17)* | *7 (5,12)* |
| 41-60 dB | 20.6(6.0) | 13.9 (4.4) | 7.8 (2.6) | 11.5 (2.3) | 7.1 (2.4) |
| | *21 (11,32)* | *12 (10,26)* | *6 (6,16)* | *11 (7,18)* | *6 (5,13)* |
| >60 dB | 24.4 (6.5) | 16.9 (6.9) | 8.6 (3.0) | 12.0 (3.0) | 6.8 (2.1) |
| | *25 (13,33)* | *15 (10,29)* | *8 (6,16)* | *12 (7,19)* | *7 (5,11)* |

Groups: normal: non-hearing impaired; unaided: hearing impaired without hearing aid; aided: hearing aid

2

The correlation coefficients between the five derived scales and objective and subjective hearing impairment are presented in Table 2.5. The *functionality* scale correlated positively with all the pure tone averages and negatively with the VAS hearing scale for the entire sample. These relations were also present within the unaided and aided groups but only held for the VAS scale in the normal hearing group. The *social hearing* scale was positively correlated with the pure tone averages for the whole sample and within the aided group. *The social hearing* scale correlated negatively with the VAS-hearing scale for the whole sample and within the aided and unaided groups. The *aid stigma* scale was only negatively correlated with pure tone averages within the normal group. Further investigation revealed that within the unaided group this relation was observed only for those persons in the WHO classification of no impairment. Examination of the unaided group relative to the WHO classification revealed that within the WHO defined mildly impaired range the correlations were stronger being 0.498 with BEPTA 1, 2, 4 kHz. *Social pressure* was positively correlated with pure tone averages and negatively correlated with the VAS scale for the entire sample, and only correlated (negatively) with the VAS scale within the unaided group. *Aid unwanted* did not appear to be related to either pure tone averages or the VAS hearing scale.

Table 2.5   Correlations scale sum scores with objective (BEPTA) and subjective hearing (VAS).

| Factor | Hearing outcome | Entire sample | Normal | Unaided | Aided |
|---|---|---|---|---|---|
| Functionality | BEPTA1,2,4kHz | 0.569** | 0.011 | 0.389** | 0.288* |
| | BEPTA 0.5,1,2,4kHz | 0.563** | 0.088 | .362** | 0.277* |
| | WEPTA 0.5,1,2,4kHz | 0.559** | .174 | .336** | 0.275* |
| | Asymmetry | -.109 | -.157 | -.072 | 0.059 |
| | VAS | -.611** | -.371** | -.481** | -.453** |
| Social hearing | BEPTA1,2,4kHz | 0.258** | -.149 | 0.139 | 0.426** |
| | BEPTA 0.5,1,2,4kHz | 0.262** | -.097 | 0.114 | 0.411** |
| | WEPTA 0.51,2,4kHz | 0.203** | -.156 | 0.076 | 0.353** |
| | Asymmetry | 0.090 | .126 | 0.000 | 0.188 |
| | VAS | -.280** | 0.005 | -.308* | -.383** |
| Aid stigma | BEPTA1,2,4kHz | 0.001 | -.311** | 0.144 | 0.233 |
| | BEPTA 0.5,1,2,4kHz | -.018 | -.271* | 0.052 | 0.178 |
| | WEPTA 0.51,2,4kHz | -.032 | -.238* | -.029 | 0.248 |
| | Asymmetry | 0.099 | 0.142 | 0.129 | 0.021 |
| | VAS | 0.009 | 0.099 | 0.018 | -.185 |
| Social pressure | BEPTA1,2,4kHz | 0.501** | 0.027 | 0.223 | 0.042 |
| | BEPTA 0.5,1,2,4kHz | 0.481** | .118 | 0.221 | -.018 |
| | WEPTA 0.51,2,4kHz | 0.482 ** | .110 | 0.252 | -.042 |
| | Asymmetry | -.111 | -.015 | -.092 | 0.014 |
| | VAS | -.414** | -.147 | -.347** | 0.074 |
| Aid unwanted | BEPTA1,2,4kHz | -.086 | .045 | 0.184 | 0.076 |
| | BEPTA 0.5,1,2,4kHz | -.103 | .022 | 0.152 | 0.015 |
| | WEPTA 0.5,1,2,4kHz | -.094 | .118 | 0.118 | 0.024 |
| | Asymmetry | 0.020 | -.149 | -.086 | 0.137 |
| | VAS | 0.050 | -.028 | -.073 | -.125 |

*P*: * < .05; **<.005

## Discussion

The purpose of this study was to validate the HARQ for use in a more heterogeneous group than for which it was originally designed and to investigate its usefulness to assess hearing dimensions in a potential screening population. Factors have been discerned which address issues prominent in the patient journey towards hearing rehabilitation.

Specifically scales were derived which relate to how the individual experiences his own hearing ability. Laplante-Levesque et al.[27] ascertained in their qualitative study that it is experienced difficulties in hearing which influence help seeking rather than clinical services, and persons who are aware of the gradual process of becoming hearing impaired feel that clinicians failed to recognize this process. In a study of 63 hearing impaired persons conducted at a geriatric outpatient clinic, Wu et al.[28] found that willingness to use a hearing aid was related to hearing functionality but not to audiometric outcomes. The awareness stage described by Manchaiah et al.[7] is marked by experienced functional and social limitations. The transition from awareness to movement will be moderated by attitudes towards hearing aids and mediated by pressure experienced to be assessed. The scales developed here could help mark the position on the patient journey with as destination successful hearing rehabilitation and resolution.

Clearly, each scale evaluates an aspect of hearing or factors related to the passage of the hearing impaired person from the (pre)-awareness to the movement stage. The *functionality* and *social hearing* scales could be applied to assess experienced hearing difficulties as part of a screen for referral for rehabilitation. In this context it might be important to realize that hearing aid fitting does not necessarily have to be the first step in the rehabilitation process keeping in mind the implementation of the Active Communication Education program by Hickson et al.[14] whereby individuals could discover the possibility of improving communication through hearing strategies.

The *functionality* items are clearly an indication of disability as these items address the limitations experienced in functional hearing. This is indicated by the fact that the mean scores obtained were lowest for the normal and highest for the aided group and further varied relative to the severity of the WHO classification. The same trend was observed for the *social hearing* scale relative to our hearing groups. However the increase in mean scores did not correspond uniformly to the severity of the WHO classification with the range for 20-40 dB having the lowest values. Perhaps the relation is compromised by Dutch policy providing hearing aid reimbursement for hearing impairment for 35 dB or more in the best ear. Nevertheless the scale appears to measure experienced social limitations and how the person feels his/her hearing limitation is perceived by others. Therefore the *functionality* and *social hearing* scales represent positions in the pre-awareness and awareness stage.

**2**

In the Active Communication Education program,[14] the greatest improvement was found among participants with higher pre-test scores for the HARQ hearing items, further supporting the notion that these HARQ items help define the patient's location in these first stages on the journey towards rehabilitation. It is thus the experienced hearing, which is at the core of awareness of potential transition to the movement stage.

The hearing aid scale *social pressure* reflects whether the individual has experienced others noticing his/her hearing problems. Even though it is composed of hearing aid items it is about hearing functionality as viewed by others. This scale may help identify persons in the pre-awareness stage where others notice the hearing limitations of the hearing impaired individual while the hearing-impaired individual does not. The other two hearing aid scales, *aid stigma* and *aid unwanted*, quantify barriers to reaching the movement stage.

The factor structure obtained for the hearing aid items is markedly different from those obtained in the original HARQ study. The only clear similarity lies in that six of their hearing aid stigma items comprise our *aid stigma* scale. One item: "Many people don't know how to treat you when you have a hearing aid" gave problems in the calculation of the polychoric matrix. This may be attributable to the fact that our sample consists of persons with and without the experience of wearing a hearing aid. Two other original HARQ Hearing Aid Stigma items had low loadings on our *aid stigma* scale. These items were: "I think the behind-the-ear aids are really quite small and inconspicuous" and "From what I know, hearing aids don't help a great deal". These items also had relatively the lowest loadings in the original study.[18] All three items from the original HARQ Pressure to be Assessed scale had high loadings on the *social pressure* scale. At the same time, however, two items from the original HARQ Aid Not Wanted scale: "I don't consider it important to be assessed for a hearing aid" and "My hearing is not so bad that I need a hearing aid" had negative loadings on the *social pressure* scale indicating either an absence or refusal to acknowledge pressure to be assessed.

In the present study, the items from the original HARQ Positive Expectation scale were dropped from the analysis because they gave problems when estimating the polychoric correlation coefficients. These items had high percentages of unusable responses. This makes sense as the content was specifically formulated for persons in the process of considering hearing aid uptake. These items also lack appropriateness for the hearing aid users in our sample in that they are experienced hearing aid users and therefore responses would be based on experience rather than expectation. Hallam and Brooks[18] saw this scale as an over-estimation of the positive effects of a hearing aid. The fact that our aided group scored slightly lower for this original HARQ scale than the other two groups reflects having experienced adjustment to using a hearing aid. Nevertheless, given the results reported by Jerram and Purdy [19] this aspect of hearing rehabilitation should be included in a screening setting. Of the three items in this scale there is only one item, which does not contain a reference to time for adjustment to

hearing aid use. This item: "I expect to hear as easily with a hearing aid as I did before" might be appropriate for screening purposes. The *aid stigma* scale appears to measure stigma, and the fact that it correlates negatively with hearing impairment only in the non-hearing impaired group may indicate acceptance, there being persons in this group approaching a level of hearing impairment warranting hearing rehabilitation. This relation between hearing impairment and a reduction in negative attitudes towards hearing aids was further supported by the fact that this relation was strongest within the normal group with no impairment and within the unaided group with moderate impairment according to the WHO classification. This notion appears to be further supported by the findings of Jerram and Purdy[19] who found no relation between HARQ hearing aid stigma scores and hearing aid outcome, suggesting that hearing aid stigma is less manifest when further along the patient journey. Our findings may also suggest that increasing hearing impairment may reduce hearing aid stigma in some individuals allowing them to benefit from hearing aid fitting at relatively lower levels of hearing impairment.

The *aid unwanted* scale is a measure of reluctance and thus a barrier to hearing aid uptake, and this is underlined by the fact that within the hearing impaired but unaided group, the *aid unwanted* scores were highest. There was also a high correlation between the VAS hearing scale and the *social pressure* scale within this group, indicating that persons scoring high on this scale may receive relatively more suggestions from others regarding their hearing ability. *Social pressure* may then be a motivator for transition into Manchaiah's[7] movement stage.

Hallam and Brooks[18] established the factor structure of the HARQ with a sample of 164 first-time hearing aid candidates ranging in age from 36-96 years (mean=74.2, SD=10.4), with 92% of the sample being age 60 or more. Our study group has a similar age distribution even though we only included persons aged 55 and more but differs regarding hearing (aid) status as it was our intent to mimic a screening population. However the HARQ items were originally formulated for persons seeking hearing rehabilitation and thus those persons who are experiencing difficulties in hearing. Thus some HARQ items are out of context for persons not experiencing those difficulties. This is precisely what Hyde[17] was warning when he emphasized the need to examine existing measures relative to purpose, context and responsiveness.

The HARQ was developed as a supplementary tool in the hearing aid rehabilitation process. It is then not surprising that a somewhat different factor structure emerged when applied to individuals not seeking rehabilitation. This result should be considered a positive finding since the objective of the present study is to determine if there are items which can be applied in a heterogeneous sample, rendering usefulness in a screening process which will by the very nature of screening include persons with normal hearing or, in any case, not consciously experiencing hearing difficulties such as persons in Manchaiah's[7] pre-awareness stage of the patient journey.

2

The inclusion of hearing aid users in this study could be questioned given that the objective is to develop tools for assessment in screening. The fifth of the seven stages of Manchaiah's patient journey[7] is rehabilitation. Thereafter follow the stages evaluation and resolution, the latter being the stage of satisfied hearing aid use. Gianopoulos et al.[29] performed a follow up of 166 persons who had been fitted with a hearing aid as a result of a hearing screen and found that only 43% were still using it after 8-16 years. Through interviews it was determined that 47 of the 66 persons who were not using their aids were still willing to try a new aid, indicating that hearing screening should be extended to unsuccessfully rehabilitated hearing aid owners.

It is extremely relevant how hearing impairment is defined. At the time of our study Dutch policy provided reimbursement for hearing aid fitting when the pure tone average for the frequencies 1, 2 and 4 Hz exceeded 35 dB in the better ear. When the WHO classification is applied our sample results in 83 persons in the category *slight hearing impairment* of which 34 are in our *unaided* and 32 in our *aided* group. The scale scores either increased or decreased monotonically relative to the WHO hearing impairment classification four scales providing validity. The only exception was the *aid stigma* scale, which did not appear to vary monotonically according to classification. Nevertheless for some ranges of hearing impairment there is a relation indicating that hearing aid stigma as a barrier to hearing rehabilitation may be moderated by increasing hearing impairment.

It is interesting to remark that the percentage of persons with asymmetry exceeding 10 dB was much greater in the hearing aid group suggesting that asymmetric hearing impairment may trigger help-seeking behaviour.

The cross-sectional design of the present study hinders making conclusive remarks about the use of the HARQ to assess the individual's progress along the patient journey. A longitudinal study would provide information as to how these various aspects of hearing experience and willingness to start hearing rehabilitation interact and influence the individual. In a longitudinal study, the use of the SADL and ECHO are to be considered. Although we had included the positive expectation scale items of the HARQ, they could not be used to define a scale and thus were not included in the final analysis. The SADL and ECHO questionnaires include more items facilitating quantification of this domain which would be particularly useful in a longitudinal study where these questionnaire responses can be compared before and after hearing aid fitting.

At the same time it should be remarked that answering a questionnaire regarding hearing and hearing aids, is in itself an intervention as it may encourage the individual to take action or at least to be more receptive to action in the direction of hearing rehabilitation.

## Conclusions

The process of becoming hearing impaired is often a gradual one. We have shown here that hearing functionality and social limitations can be assessed with elements of the HARQ. Furthermore, barriers to hearing uptake can be quantified with items addressing pressure to be assessed and hearing aid stigma. Thus clearly these derived scales would supplement audiological assessment in hearing screens, to assess receptivity of hearing impaired individuals to the process of rehabilitation and identify an individual's location on the patient journey. In particular, the scales developed here may provide more clues to understand the relatively unknown early stages of the hearing impaired patient journey.

2

# References

1. Mulrow CD, Aguilar C, Endicott JE, Velez R, Tuley MR, Charlip WS, Rhodes MC, Hill JA, De Nino LA. Association between hearing impairment and the quality of life of elderly individuals. *J Am Geriatrs Soc* 1990;38:45-50.

2. Appollonio I, Carabellese C, Frattola L, Trabucchi M. Effects of sensory aids on the quality of life and mortality of elderly people: a multivariate analysis. *Age Ageing* 1996;25:89-96.

3. Joore MA, Van der Stel H, Peters HJM, Boas GM, Anteunis LJC. The cost-effectiveness of hearing aid Fitting in the Netherlands. *Arch Otolaryngol Head Neck Surg* 2003;129:297-304.

4. Hallberg L, Carlsson S. Hearing impairment, coping and perceived hearing handicap in middle-aged subjects with acquired hearing loss. *Br J Audiol* 1991;25.323-330.

5. Hetu R. (1996). The stigma attached to hearing impairment. *Scan Audiol* 1996;43:2-24.

6. Lunner T. Cognitive function in relation to hearing aid use. *Int J Audiol* 2003;42:S49-S58.

7. Manchaiah VKS, Stephens D, Meredith R. The patient journey of adults with hearing impairment: the patients' views. *Clin Otolaryngol* 2011;36:227-234.

8. Yueh B, Collins MP, Souza PE, Boyko EJ, Loovis CF, Heagerty PJ, Liu CF, Hedrick SC. Long-term effectiveness of screening for hearing loss: the screening for auditory impairment – Which hearing assessment test (SAI-WHAT) Randomized Trial. *J Am Geriatr Soc* 2010;58:427-434.

9. Stephens D, Kramer SE. Living with hearing difficulties: The process of enablement. Wile, Chichester; 2009.

10. Knudsen LV, Oberg M, Nielsen C, Naylor G, Kramer SE. Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids: A review of the literature. *Trends Amplif* 2010;14:127-154.

11. Yueh B, Shapiro N, MacLean CH, Shekelle PG. Screening and Management of Adult Hearing Loss in Primary Care. *JAMA* 2003;289:1976-1885.

12. Ventry IM, Weinstein BV. The hearing handicap inventory for the elderly: a new tool. *Ear Hear* 1982;3:128-134.

13. Pronk M, Kramer SE, Davis AC, Stephens D, Parazzini M, Smith PA, Thodi C, Anteunis LJC, Grandori F. Interventions following hearing screening in adults: A systematic descriptive review. *Int J Audiol* 2011;50:594-609.

14. Hickson L, Worrall L, Scarinci N. A randomized controlled trial evaluating the active communication education program for older people with hearing impairment. *Ear Hear* 2007;28:212-230.

15. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psych Bull* 1995;52:281-302.

16. Terwee CB, Bot SDM, De Boer MR, Van der Windt DAWM, Knol DL, Dekker J, Bouter LM, De Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.

17. Hyde ML. Reasonable psychometric standards for self-report outcome measures in audiological rehabilitation. *Ear Hear* 2000;21:24-36.

18. Hallam RS, Brooks DN. Development of the Hearing Attitudes in Rehabilitation Questionnaire (HARQ). *Br J Audiol* 1996;30:199-213.

19. Jerram JCK, Purdy SC. Technology, expectations, and adjustment to hearing loss: predictors of hearing aid outcome. *J Am Acad Audiol* 2001;10:445-457.

20. Meister H, Walger M, Brehmer D, Von Wedel U-C, Von Wedel H. The relationship between pre-fitting expectations and willingness to use hearing aids. *Int J Audiol* 2008;47:53-159.

21. Cox RM, Alexander GC. Measuring Satisfaction with Amplification: The SADL scale. *Ear Hear* 1999;20: 306-320.

22. Cox RM, Alexander GC. Expectations about hearing aids and their relationship to fitting outcome. *J Am Acad Audiol* 2000;11: 68-382.23.

23. Holgado-Tello F, Chacon-Moscoso S, Barbero-Garcia I, Vila-Abad E. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Qual Quan* 2010;44: 53-166.

24. Jolliffe IT. Principal Component Analysis New York, N.Y: Springer-Verlag, 1986:96-97.

25. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika* 1965;30: 179-185.
26. Field F. Discovering Statistics Using SPSS. 3rd ed.  Sage Publications. London. Chapter 17, 2005.
27. Laplante-Lévesque A,  Knudsen LV, Preminger JE, Jones L, Nielsen C, Oberg M, Lunner T, Hickson L, Naylor G, Kramer SE. Hearing help-seeking and rehabilitation: Perspectives of adults with hearing impairment. *Int J Audiol* 2012;51:91-102.
28. Wu H, Chin J, Tong H. Screening for hearing impairment in a cohort of elderly patients attending a hospital geriatric medicine service. *Singapore Med J* 2004;45:79-84.25.
29. Gianopoulos I, Stephens D, Davis A. Follow up of people fitted with hearing aids after adult hearing screening: the need for support after fitting. *BMJ* 2002;325:471.

2

# Chapter 3

## Quantification of experienced hearing problems with Item Response Theory

MN Chenault
MPF Berger
B Kremer
LJC Anteunis

# Abstract

3

**Purpose**

To improve the effectiveness of adult hearing screens and interventions assessment methods are needed which address the individual's experienced hearing. Item Response Theory, which provides a methodology for assessing Patient Reported Outcomes, will be examined here to demonstrate its usefulness in hearing screens and interventions.

**Method**

The Graded Response Model is applied to a scale of 11 items assessing perceived hearing functioning and 10 items assessing experienced social limitations completed by a sample of 212 persons aged 55+. Fixed and variable slope models are compared. Discrimination and threshold parameters are estimated and information functions evaluated.

**Results**

Variable slope models for both scales provided the best fit. The estimated discrimination parameters for all items except for one in each scale were good if not excellent (1.5–3.4). Threshold values varied demonstrating the complementary and supplementary value of items within a scale. The information provided by each item varies relative to trait values so that each scale of items provides information over a wider range of trait values.

**Conclusion**

IRT methodology facilitates the comparison of items relative to their discriminative ability and information provided and thus provides a basis for the selection of items for application in a screening setting.

## Introduction

A description of the hearing impaired patient's journey towards hearing rehabilitation in order to improve hearing screens and interventions is needed.[1] Individual hearing experience can be assessed with questionnaire items also known as Patient Reported Outcomes (PRO's). The past decade has witnessed a growth in the application of PRO's reflecting an increase in attention for patient outcomes which cannot be assessed directly. The settings within which PRO's are applied vary greatly with regard to both health outcomes and objectives. The object of measurement then varies with outcomes such as experienced pain, physical functional limitations, health related quality of life, and patient satisfaction with health care. Such outcomes cannot be directly measured and are therefore referred to as latent traits.

There are two general approaches to quantifying such latent traits, namely Classical Test Theory (CTT) and Item Response Theory (IRT). Traditionally, CTT has been applied whereby items comprising a scale measuring a certain trait are simply summed, ignoring trait related variation, to obtain a score indicating the strength of the trait. In contrast, IRT models the probability of a response to an item as a function of the underlying trait. The CTT approach lacks the precision afforded by IRT where the relative information provided by each item in a scale is evaluated. Additionally, as pointed out by Reeve et al.,[2] the CTT approach often results in floor and ceiling effects restricting the comparison of different severity or functional levels. Health researchers have therefore been turning towards IRT methodology in the application of PRO's. The increasing use of IRT methodology creates the need for clear standards in its application.[3,4] It has been argued, however, that both CTT and IRT methodology should be applied as complementary tools in evaluating health outcomes.[5]

In IRT methodology the focus is on the items which comprise the PRO, whereby the information provided by each item in a scale as well as the whole scale is examined. It facilitates a more thorough and useful quantification of underlying traits and/or attitudes than does CTT. Questionnaire items vary in content and specificity which is reflected by the obtained response distribution. Some items display more discriminative power than others and furthermore this discriminative power varies relative to the strength of the latent trait being quantified. A number of studies have been directed towards examining the relation between CTT and IRT, in particular how the estimation processes are related to each other and under some conditions equivalent.[6,7] The theoretical objectives of the two approaches may, however, differ.

At the same time there is a discussion going on regarding adult hearing screening and the potential role it could adopt in hearing rehabilitation. While evidence has been provided that earlier intervention means greater gain in effective rehabilitation,[8] there are barriers to rehabilitation most notably in the individual's (lack of) awareness of his/her own hearing impairment. Manchaiah et al.[9] have described the patient journey towards hearing rehabilitation as having seven stages whereby the first two, "pre-

3

awareness" and "awareness", are the most difficult to assess. Objective hearing assessment may determine whether a criterion for hearing aid reimbursement is met. But it is the experienced hearing problems which spur the individual to seek help and form the base for willingness to rehabilitate. Factors related to hearing problems experienced must therefore be examined, calling for questionnaire items addressing the subjective hearing experience, and specific contexts within which hearing difficulties occur.

IRT methodology has been applied in the field of audiology and to some extent in a screening context. Demorest et al.[10] applied IRT methodology to the Communication Profile for the Hearing Impaired (CPHI) to obtain two scales: one addressing communication and the other adjustment aspects of hearing impairment. In their study, the CPHI items, which are on a 5 point scale, were dichotomized to obtain pass-fail outcomes for application in hearing screening to determine rehabilitation needs. Mokkink et al.[11] also applied IRT to improve the quality and application of a Dutch version of the CPHI.

In the present study IRT modelling will be applied to scales obtained through psychometric evaluation of the hearing items of the Hearing Aid Rehabilitation Questionnaire (HARQ) which was developed by Hallam and Brooks.[12] Two scales pertaining to how the individual's experienced hearing were discerned: one scale addresses perceived functional limitations and the other experienced limitations in social settings.[13]

The main goal of this present study is to apply IRT methodology to these two hearing scales. The two scales and their individual items will be evaluated within the IRT methodology framework keeping in mind their usefulness in a hearing screen setting. Ideally the calibration of these scales will facilitate the quantification of the stage at which a person is on the patient journey towards hearing rehabilitation.

## Methodology

### Data sample

The item responses considered in this study were obtained by administering the HARQ to 212 Dutch persons aged 55 and older consisting of 63 hearing aid users and 149 non-users with a mean pure tone average best ear (PTABE for 1, 2, 4 kHz) of 38 dB (s.d.16). The hearing aid users were assessed without their hearing aid. Hearing aid users were included since a potential screening population will also include persons who have been fitted with a hearing aid and either do or do not use it.

### HARQ

The hearing scales considered here were obtained by applying exploratory factor analysis to 20 hearing items of the HARQ whereby the polychoric correlation matrix of

the 212 questionnaire responses were analysed. The same factor structure was obtained through orthogonal and non-orthogonal rotations. The two hearing scales obtained were labelled *functionality* (perceived hearing functionality) and *social hearing* (experienced hearing related social limitations). The *functionality* scale consists of 11 items, while the *social hearing* scale consists of 10 items with all items being on a 3-point ordinal response scale: "agree", "partly agree", and "disagree". Two items loaded on both scales. Reliability analysis resulted in a Cronbach's alpha of 0.85 for the *functionality* scale and 0.84 for the *social hearing* scale. The content of the items comprising these two scales, their non-orthogonal factor loadings, as well as explained variance and reliability of the two scales are presented in Appendix I. The component correlation was -.331. The response categories of 4 items of the *functionality* scale were recoded so that the direction of the trait being measured conformed to the direction of the other 7 items in the scale.

## Testing assumptions necessary for IRT estimation

Monotonicity will be confirmed by checking whether the individual item responses are non-monotone decreasing, relative to the trait scores obtained through IRT estimation. Uni-dimensionality will be evaluated with confirmatory factor analysis (CFA) according to the methodology proposed by Joreskog,[13] whereby the item coefficients, which are statistically significantly different from zero, and the goodness of fit statistics, Root Mean Square Error of Approximation (RMSEA), the Non-Normed Fit Index (NNFI and Comparative Fit Index (CFI), will be considered. Local independence will be evaluated by examining the residual correlations of the items within their scale.

## IRT estimation

Samejima's[15,16] Graded Response Model (GRM) will be applied because the responses are on a three-point ordinal scale of "agree", "partly agree" and "disagree". The fit of two models will be compared: one model with a single slope for all items and one model with varying slopes per item. The chi-square statistic of the differences between the two models will be the outcome considered, whereby a high chi-square statistic indicates that a variable slope model has a better fit than a fixed slope model. The estimated item parameters and their respective ICC curves will then be examined and compared regarding their relative discrimination and threshold parameters. Thereafter the information curves of the items of each scale and the total information curve will be examined. A more detailed explanation of IRT related methodology is given in Appendix II.

## Comparison of the three category response model to two category response model

To investigate the usefulness of the middle response category "partly agree", models will also be obtained whereby the two response categories "agree" and "partly agree" are merged so that there are two response categories left: "agree" and "disagree".

## Data preparation and analysis

Lisrel 8.7 will be applied to assess assumptions for IRT.[17] Multilog 7.3 will be employed to obtain GRM model parameters.[18]

# Results

## Monotonicity, uni-dimensionality, local independence

The mean trait scores per response category were non-decreasing relative to the lower categories. The RMSEA's for the *functionality* and *social hearing* scales were acceptable at 0.0791 and 0.0843, respectively, although values less than 0.08 are preferred. The NNFI's and CFI's were 0.96 and 0.97 for the *functionality* and 0.96 and 0.97 for the *social hearing* scales, with all item loading coefficients being highly statistically significant, providing strong evidence of uni-dimensionality for each scale (Table 3.1). The residual correlations were 0.195 or less for the *functionality* scale and 0.170 or less for the *social hearing* scale indicating local independence.

Table 3.1    Confirmatory factor analysis (CFA) for uni-dimentionality assumption (all *t*-statistics were highly significant with *p*<.0005)

| Items | | CFA loading (s.e.) | *t*-test |
|---|---|---|---|
| *Functionality* (RMSEA= .0791, NNFI= 0.964, CFI= 0.971) | f1 | 1.86 (0.101) | 18.43 |
| | f2 | 3.92 (0.237) | 16.55 |
| | f3 | 1.81 (0.131) | 13.73 |
| | f4 | 3.83 (0.112) | 34.21 |
| | f5 | 2.69 (0.059) | 45.83 |
| | f6 | 2.23 (0.082) | 27.78 |
| | f7 | 3.12 (0.173) | 18.03 |
| | f8 | 1.63 (0.047) | 35.10 |
| | f9 | 2.32 (0.044) | 52.74 |
| | f10 | 2.30 (0.085) | 27.12 |
| | f11 | 4.02 (0.185) | 21.77 |
| *Social hearing* (RMSEA= .0843, NNFI= 0.964, CFI= 0.972) | s1 | 1.058 (0.037) | 28.27 |
| | s2 | 0.892 (0.046) | 19.46 |
| | s3 | 2.547 (0.130) | 19.58 |
| | s4 | 0.786 (0.061) | 12.85 |
| | s5 | 0.978 (0.024) | 40.55 |
| | s6 | 0.874 (0.039) | 22.31 |
| | s7 | 1.379 (0.095) | 14.52 |
| | s8 | 0.857 (0.027) | 32.39 |
| | s9 | 1.363 (0.030) | 45.65 |
| | s10 | 1.284 (0.044) | 29.19 |

## Comparison of single slope and varying slope GRM models

For the *functionality* scale the variable slope model with slopes varying per item had a statistically better fit than the single slope model, the chi-square=49.1 (df=10, *p*=.0000) for the difference between the two models. The discrimination parameter was estimated at 1.77 (s.e.=0.08) for the single slope model while it ranged from 0.78 to 3.39 for the items of the variable slope model, with total information peaked at 13.7.

The variable slope model for *social hearing* also had a better fit than the single slope model, with a chi-square=42.5 (df=9, *p*=.0000). Start parameter values were needed to obtain convergence and thus a solution. Various start values were applied which resulted in almost identical values so that it could be assumed that the estimates were stable. The single slope model discrimination parameter was estimated at 1.86 (s.e.=0.09) while it ranged from 0.64 to 3.33 for the variable slope model, for which total information peaked at 13.0.

The estimated parameters for the variable slope model for both scales are presented in Table 3.2. The last two columns contain the estimated parameters when "agree" and "partly agree" responses are merged, leaving just one threshold parameter to be estimated. It can be seen that the threshold parameter for the dichotomized items are almost identical to the second threshold parameter for the three response version, except for the items of the *functionality* scale which were formulated in the other direction than the other 7 items. For these four items, the threshold parameter estimations for the dichotomized items are close to those for the first threshold parameter, as would be expected.

Table 3.2 Estimated GRM model parameters for *functionality* and *social hearing* (*a*: discrimination parameter; $b_i$: threshold parameters).

| Scale items | | 3-point response | | | 2-point response | |
|---|---|---|---|---|---|---|
| | | $a$ | $b_1$ | $b_2$ | $a$ | $b_1$ |
| *functionality* | 1 | 1.86 | -.39 | .17 | 1.94 | 0.17 |
| | 2 | 2.03 | -1.73 | -1.27 | 2.06 | -1.27 |
| | 3 | 0.78 | -.37 | .52 | 0.83 | 0.53 |
| | 4 | 2.86 | -1.20 | -.87 | 3.34 | -0.85 |
| | 5 | 2.47 | -1.05 | -.59 | 2.86 | -0.56 |
| | 6 | 2.03 | -.38 | .14 | 1.80 | 0.13 |
| | 7 | 1.47 | -1.13 | -.67 | 1.45 | -0.68 |
| | 8 | 2.01 | -1.88 | -1.00 | 1.75 | -2.01 |
| | 9 | 3.39 | -1.12 | -.62 | 4.65 | -1.11 |
| | 10 | 1.46 | -.31 | .27 | 1.33 | -0.30 |
| | 11 | 1.33 | -1.17 | -.78 | 1.07 | -1.32 |
| *social hearing* | 1 | 2.37 | -1.57 | -1.12 | 2.44 | -1.12 |
| | 2 | 1.56 | -1.63 | -.97 | 1.63 | -0.95 |
| | 3 | 2.73 | -1.21 | -.85 | 2.80 | -0.86 |
| | 4 | 0.64 | -4.14 | -1.65 | 0.63 | -1.68 |
| | 5 | 2.68 | -2.09 | -1.72 | 2.51 | -1.81 |
| | 6 | 1.48 | -2.80 | -1.85 | 1.44 | -1.93 |
| | 7 | 1.54 | -2.20 | -1.34 | 1.52 | -1.38 |
| | 8 | 2.28 | -2.50 | -1.93 | 1.98 | -2.11 |
| | 9 | 3.33 | -1.35 | -1.02 | 3.44 | -1.02 |
| | 10 | 2.43 | -1.07 | -0.65 | 2.42 | -0.67 |

## Item Characteristic Curves

The Item Characteristic Curve (ICC) provides a graphical representation of an item's discriminative ability and thresholds. Figure 3.1 presents the individual ICC curves for the 11 *functionality* and the 10 *social hearing* items based on the estimated parameters in Table 3.2. Item f3 of the *functionality* and item s4 of the *social hearing* had low discrimination parameters of 0.78 and 0.64 respectively. The highest discrimination parameters, which exceed 3.3, were obtained for items f9 and s9 of the *functionality*, respectively *social hearing* scales.

Items f2 and f8 of the *functionality* and items s4 and s6 of the social *hearing* scale had relatively lower first thresholds. These lower threshold values indicate that these items are more likely to elicit a "partly agree" relative to an "agree" response than the other items in the scale.
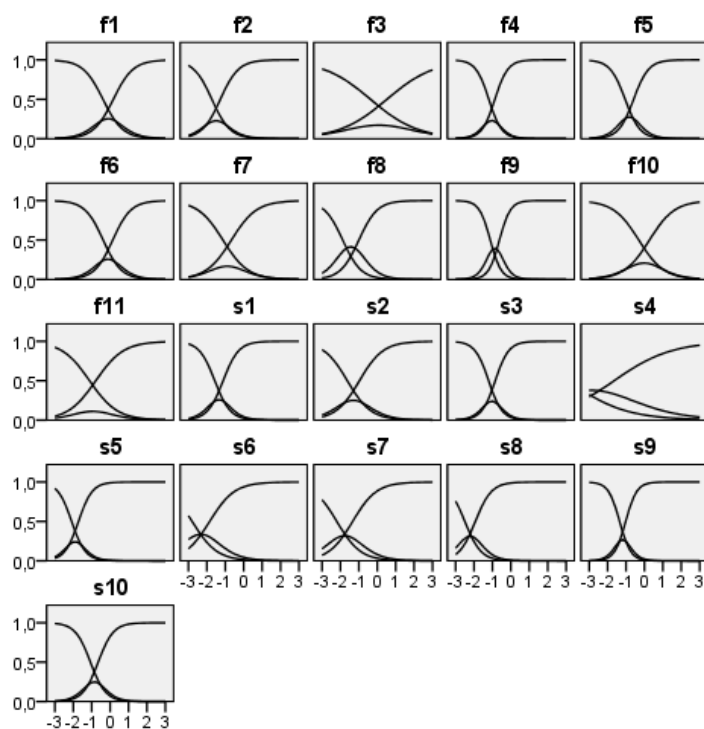


Figure 3.1    ICC curves of 11 functionality items (fi) and 10 social hearing items (si).

## Information curves

An information curve displays the amount of information provided by each item relative to trait values. In Figure 3.2 the *functionality* and *social hearing* item

information curves are presented. *Functionality* item f3 and *social hearing* item s4, which had unacceptably low discrimination parameters, have the flattest information curves. The information curves of *functionality* items f4, f5 and f9 and *social hearing* items s3, s5 and s9 have relatively higher peaks than the other items, indicating relatively higher information content within certain ranges of trait values.
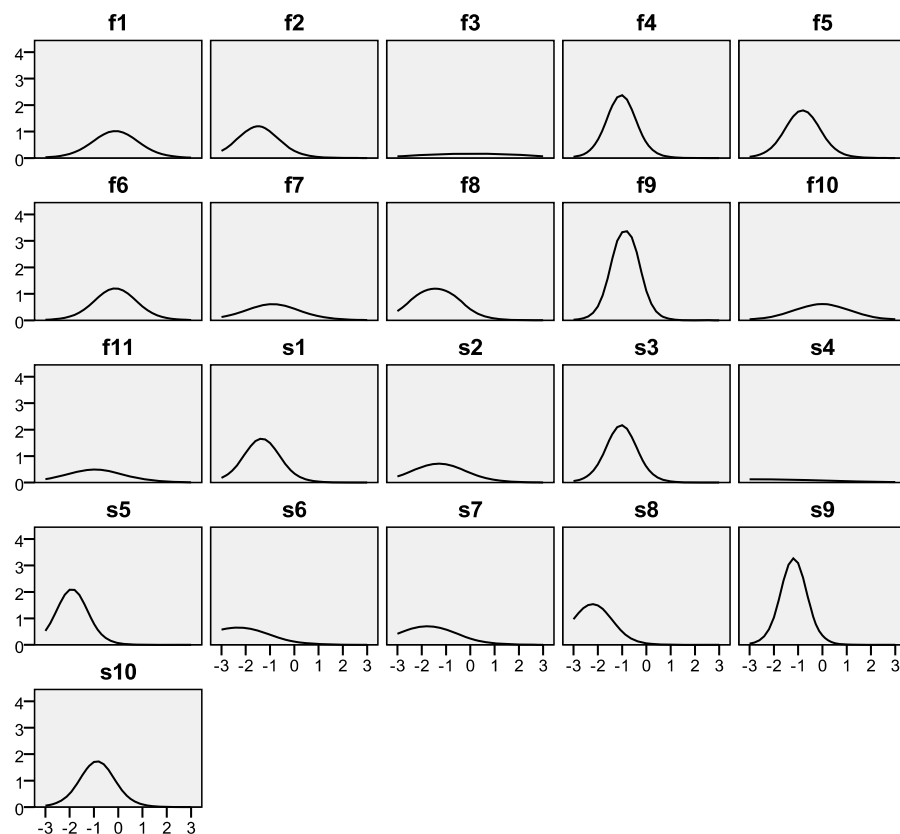


Figure 3.2     Information curves of 11 functionality items (fi) and 10 social hearing items (si).

## Information curves for 3-point versus 2-point response categories

In Figure 3.3 the total information curves for three versus two response categories are compared. The information curves of the 3-point response and the 2-point response of the *functionality* scale are almost identical, suggesting that the middle category provides marginal information. In contrast, the information curve for the 3-point *social hearing* scale clearly provides more information than the 2-point scale, since the total information curve is not only taller but also wider. Contributing to this difference

between the two scales is the fact that 4 of the 11 items of the *functionality* scale in the original scale were phrased so that an "agree" response corresponds to a "disagree" for the first 7 items in the scale, so that the wording of these items is mirrored. The bottom graph of Figure 3.3 compares the 3-point and 2-point information curves for the 7 items of the *functionality* scale which were not mirrored (thus omitting the last 4 mirrored items of this scale). The same differences are observed between the curves obtained for the 3-point and 2-point responses for the *social hearing* scale, namely that the left slope of the information curves is shifted slightly to the right.
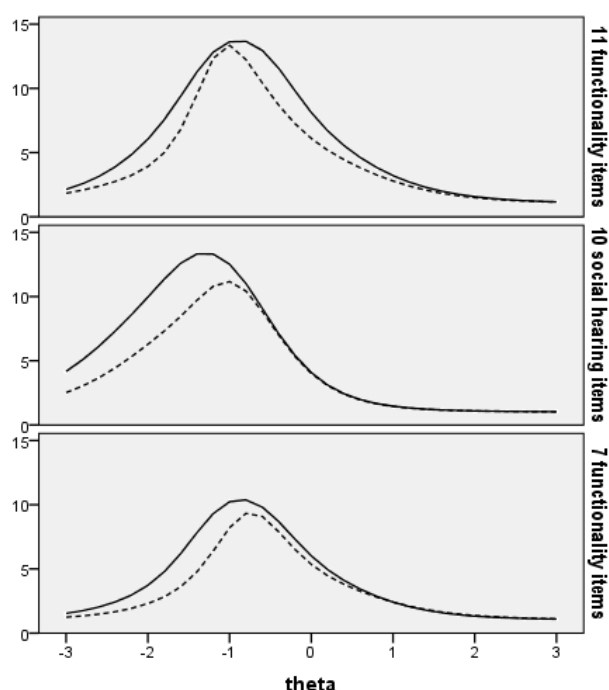


Figure 3.3    Comparing information curves for 3 (solid line) versus 2 (dotted line) response categories for the functionality and social hearing scales.

## Discussion and conclusions

This paper presents the application of IRT to items related to hearing. The objective is to develop tools to support traditional hearing assessment in the hearing screen setting. Many authors have discussed the fact that it is not the level of objectively measured hearing impairment but rather experienced hearing difficulties which spurn help-seeking and initiation of hearing rehabilitation. Here we have applied IRT models to quantify experienced hearing difficulties not only to determine which items are more

discriminatory, but also which items offer relatively more or less information for lower or higher values of the underlying trait or hearing experience which is being quantified. Classical test theory has been the most widely used approach to investigate the measurement attributes of questionnaire items. In the classical approach sum scores quantifying the underlying trait are obtained with all items contributing equally, while in IRT methodology the relative differences in items are considered. Not only does the discriminative ability vary per item but also the amount of information provided. The fact that items vary in the amount of information they provide along the trait continuum illustrates both their supplementary and complementary roles.

*Functionality* item f3 had a very poor discrimination value of 0.78. Examination of the content of the item: "I think I've overcome any hearing difficulties I might have through my own efforts" provides an explanation for this in that it targets how the individual copes with a supposed hearing problem. *Functionality* item f9 "By and large I am able to hear without difficulty" had the highest discrimination parameter and it is the most directly formulated question regarding how one rates one's hearing functionality.

*Functionality* items f7, f10 and f11 had sufficiently high discrimination parameters but these were relatively lower than the other items in the scale. The content of these items may provide a clue for this. Item f7 has an emotional content while items f10 and f11 may reflect the presence of denial among some subjects. These items are also very general questions about hearing problems or coping ability and may thus result in a less specific response pattern, given the heterogeneity in this sample with respect to hearing impairment and hearing aid use.

*Social hearing* item s4 had a very low discrimination parameter, an extremely low first threshold and provided little information. This makes sense when examining its textual content: "When you have hearing difficulties, other people ignore you". This item may be more directed towards how socially sensitive a person is rather than the person's own hearing experience.

One clear difference between the *functionality* and *social hearing* scales is that the latter scale has more items with more than 10% unusable responses. This may explain why starting values were required to obtain parameter estimations for the *social hearing* scale.

The two *social hearing* items demonstrating the greatest discriminative ability were item s9: "My hearing loss makes me feel isolated from other people" and item s3: "I find myself avoiding company because conversation is too much effort". Both of these items address hearing in a conversation setting. Their threshold values are also quite similar being respectively -1.35 and -1.02 for item s9 and -1.21 and -.85 for item s3. These are larger than those obtained for item s5: "As I see it, I am less of a person because of my hearing difficulty" indicating that for this item there is a quicker crossover from "agree" to "partly agree", which makes sense when considering the dramatic content of this item.

3

In examining the total information curves, it is clear that the items within a scale provide varying amounts of information relative to theta or trait values. Summing the information values indicates how much information the items provide together. The varying threshold parameters give rise to an aggregate of information being provided over a wider range of trait values.

The middle curve of the ICC curves for the items of both scales tend to be low relative to the height of the two outer curves, suggesting a very moderate contribution from the middle response category "partly agree". When the "agree" and "partly agree" categories were merged, there appeared to be some loss of information, which is in line with what Samejima has shown,[14] even though here the loss appears to be marginal. The merging of the categories "agree" and "partly agree" could be considered for practical reasons. In the screen setting study by Demorest et al.[10] responses to the 5-point CPHI scale items were dichotomized for conceptual reasons whereby the response categories "frequently" and "usually, almost always" formed a passing score for communication problems while the responses "about half the time", "occasionally, sometimes" and "rarely, almost never" were a fail for hearing problems. It appears then to be a matter of choosing between practical considerations and the loss of information incurred in doing so.  The main concern here would be, that by reducing the number of categories the greatest loss of information would occur in the lower theta range, where we expect persons to be located who are in the pre-awareness and awareness stages of hearing impairment, stages where screening might be most crucial.

The IRT models estimated were based on responses from a heterogeneous sample. Not only did objectively measured hearing levels range from normal to moderate hearing impairment levels but the sample also included hearing aid owners/users. There were non-applicable or don't know responses for more than 10% of the cases for some items as some of the items are formulated specifically for persons with hearing impairment. IRT utilizes the responses of all the items in the scale relative to the underlying trait being measured and in this way can quantify a person's trait in the event of missing values. Low discrimination parameters reflect the variability in response independent of the underlying trait being quantified and may indicate that the item is not interpreted by all the respondents in the same way.

Item Response Theory provides methodology which can be utilized in audiology to evaluate items to develop scales to assess the limitations the individual experiences in functioning and social situations as a result of his or her hearing impairment. In the present study, psychometric aspects of these items have been explored with IRT. The next step could be to use these IRT models to obtain individual trait scores. These scores can be used to compare to other hearing outcomes such as objective audiometric assessment. In particular it would be interesting to examine the accuracy of trait score estimates relative to the strength of the trait. Furthermore, it would be interesting to determine to what extent these IRT obtained trait scores explain hearing

aid ownership This methodology could be utilized in a prospective study to predict hearing aid uptake and successful use.

IRT methodology offers a framework within which questionnaire items addressing experienced hearing can be calibrated for use in hearing screening. It allows the investigation of the relative contribution of items according to the underlying hearing trait being examined. Further utilization of this methodology could provide an important contribution to the assessment of persons at the early stages of awareness of their hearing impairment.

3

## Appendix I

### Content of *functionality* and *social frustration* scale items with CTT results

| Items | | EFA loading |
|---|---|---|
| *Functionality* | | |
| *Explained variance 36.1%; Cronbach's alpha .85* | | |
| 1 | It sometimes depresses me when I cannot follow a conversation | .85 |
| 2 | I dread meeting new people since becoming hearing impaired | .47 |
| 3 | I think I've overcome any hearing difficulties I might have through my own efforts | .50 |
| 4* | I find myself avoiding company because conversation is too much effort | .55 |
| 5 | In a conversational group I keep quiet for fear of saying the wrong thing | .66 |
| 6 | When several people are chatting, it bothers me that I often lose the thread of the conversation | .75 |
| 7 | It really upsets me when I realise I've got the "wrong end of the stick" in conversation | .61 |
| 8 | I can hear well enough when I really concentrate *reversed* | .70 |
| 9 | By and large I am able to hear without difficulty *reversed* | .86 |
| 10 | My hearing problems are really quite minor *reversed* | .70 |
| 11 | Difficulty in hearing is not of major concern to me at the moment *reversed* | .69 |
| *Social hearing* | | |
| *Explained variance 30.8%; Cronbach's alpha .84* | | |
| 1 | I dread meeting new people since becoming hearing impaired | .59 |
| 2 | My poor hearing sometimes makes me feel really inadequate | .74 |
| 3* | I find myself avoiding company because conversation is too much effort | .58 |
| 4 | When you have hearing difficulties, other people ignore you | .67 |
| 5 | As I see it, I am less of a person because of my hearing difficulty | .65 |
| 6 | I am sure that some people think I am stupid | .71 |
| 7 | I get the feeling that other people find it a strain to talk to me | .70 |
| 8 | Some people avoid me because of my hearing ability | .75 |
| 9 | My hearing loss makes me feel isolated from other people | .74 |
| 10 | I have to admit that deep down I feel restricted by my hearing loss | .63 |

*this item loaded on both scales

## Appendix II

### Supplemental explanation of IRT modelling

#### IRT Modelling

There are two types of parameters which are estimated in IRT: discrimination and threshold. Higher discrimination parameters indicate better discriminative ability.

Higher thresholds indicate that higher trait values are required for the transition from one response category into the adjacent higher response category. Here the responses to the items of the *functionality* and *social hearing* scales are ordered so that the trait is on a scale from low to high with higher values indicating relatively less impediment in hearing functioning and hearing in social situations.

The items considered here are on a 3-point ordinal scale and therefore the IRT model proposed by Samejima called the Graded Response Model (GRM) is appropriate.[15,16] A model can be estimated whereby the discrimination parameter is fixed for all items or where it varies per item. Since there are more parameters to be estimated when the discrimination parameter is allowed to vary, the choice of the model depends on the relative fit to the observed data. In the following estimation of variable slope models will be considered.

As an example, consider a HARQ item such as: "When several people are chatting, it bothers me that I often lose the thread of the conversation". With IRT modelling we estimate the trait values of hearing functionality which mark the transition from the responses "agree" to "partly agree" and from "partly agree" to "disagree". IRT estimation furthermore determines how sharp that transition is, in other words how distinct adjacent response categories are from each other. Higher values for the discrimination or slope parameter indicate more discriminatory power, quantifying the change in response relative to trait level. In a practical setting this means that items with higher discriminatory parameters can clearly pinpoint those persons who have relatively more problems. There are no clear rules as to what values are acceptable for these discrimination or slope estimates. Baker provided cut-off points to classify discrimination parameters for an IRT model with two response categories[19]: $a<0.2$ is very low, $0.2<a<0.4$ is low discrimination, $0.4 < a < 0.8$ moderate discrimination, $0.8<a<1$ is high discrimination, and $a≥1$ is very high discrimination. In a study applying the graded response model, Flannery et al.[20] stipulated that values $a<0.8$ are unacceptable, $0.8<a<1.0$ are marginally acceptable and $a>1.0$ are acceptable.

Given that the items considered here are on a 3-point scale: "agree", "partly agree", and "disagree", there are two threshold parameters to be estimated. The two estimated threshold parameters, $b_1$ and $b_2$, vary per item. The first threshold value, $b_1$, marks the point where there is 50% chance of crossing from the lowest response category "agree" to a higher one, either "partly agree" or "disagree", while the second threshold, $b_2$, marks the 50% chance transition from the joined categories of "agree" and" "partly agree" to "disagree".

## Item Characteristic Curve

The Item Characteristic Curve (ICC) of each item provides a graphical representation of the item's discriminative ability and thresholds. In Figure 3.4 the ICC's of four *functionality* items are presented: item f3: "I think I've overcome any hearing difficulties I might have through my own efforts"; item f6: "When several people are

chatting, it bothers me that I often lose the thread of the conversation"; item f8: "I can hear well enough when I really concentrate"; and item f9: "By and large I am able to hear without difficulty". It must be noted that items f8 and f9 have been recoded reversely. Higher values reflect less hearing difficulties. The items have discrimination parameters of respectively 0.78, 2.03, 2.01 and 3.39 which in the graph is illustrated by the steepness of the curves. For item f3, the slope or discrimination parameter was much lower indicating that there is less discrimination between the response categories. The horizontal scale is the trait scale which is often symbolized by the Greek letter theta. The ICC curves vary in their placement on the horizontal or trait axis which determines the item specific estimated threshold values. The location of the threshold values of item f3, namely -0.37 for $b_1$ and 0.52 for $b_2$, are to the right of those for item f8, being -1.88 and -1.00, indicating that the transition from an "agree" response to a "partly agree" response occurs at relatively higher theta values for item f3 than for item f8.

The relative lowness of the curve for the middle response category of item f3 reflects the low response in this middle category suggesting that omission of the middle category could be considered. In our example this would mean merging response categories "agree" and "partly agree". For each item for which the categories have been merged, there is then one less threshold parameter to be estimated.
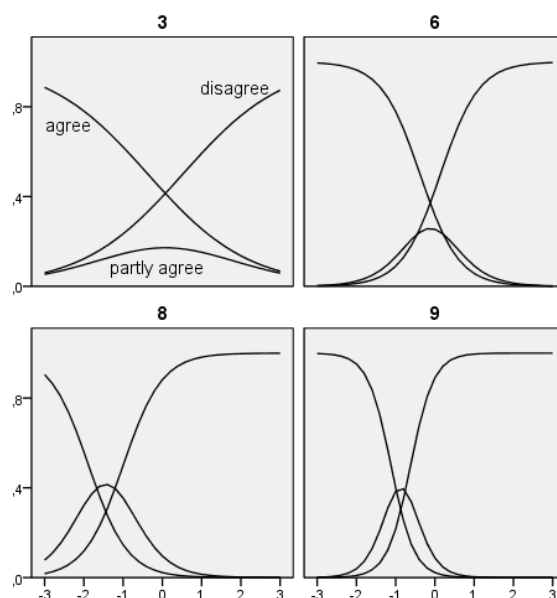


Figure 3.4    Item Characteristic Curves (ICC) for items 3, 6, 8 and 9 of the functionality scale.

## Information

The item response model, as specified by the discrimination and threshold parameters, results in an item response function which describes the probability of response relative to the trait being quantified. This relation can be transformed into an item information curve. The information curve plots the amount of information provided by each item relative to trait values. Information curves can range in form from being very peaked for a narrow range to being rather flat over a wide range of theta (trait) values. For diagnostic purposes a peaked information curve is more desirable. When the goal is to classify patients on a wider continuum, a peaked information curve may be less useful than one where the information is relatively constant over a wide range.[21] The items summed within a scale generate the total information curve. In Figure 3.5 two sets of information curves are presented. On the left the information curves of items f3, f6 and f8 and their total are shown. It can be seen that item f3 provides very little information over the trait continuum. Item f6 provides relatively more information for higher and item f8 for lower theta values. The items differ then in the trait level at which information is at its maximum. The result of including item f9, whereby the total information curve is now more peaked with a higher maximum value, can be seen in the right part of Figure 3.5. A thorough and clear explanation of the information curve as given by Baker can be found online.[19]
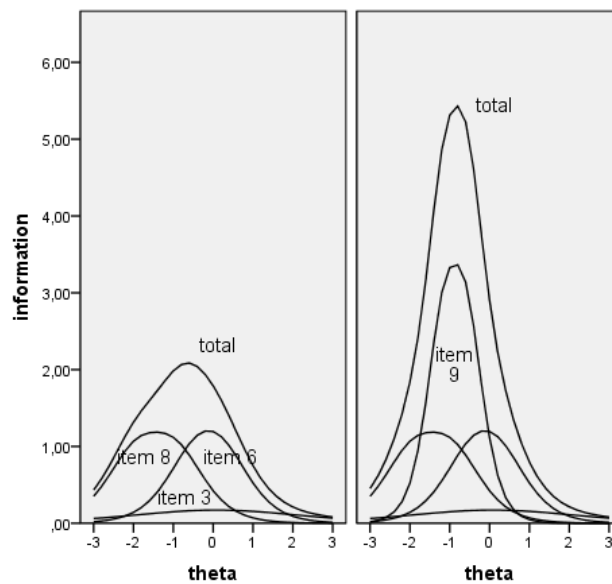


Figure 3.5    Individual and total information curves for 3 and 4 items of the functionality scale.

## Assumptions for IRT models

There are a number of assumptions made when performing IRT modelling. These include monotonicity, measurement equivalence, local independence and uni-dimensionality. The monotonicity assumption means that with an increase in the latent trait being measured, the probability of endorsing one response category higher is non-decreasing.

Measurement equivalence is concerned with the parameter estimations being invariant across subgroups. When the parameters vary across subgroups we speak of differential item functioning. This has been of particular importance in educational testing when due differences in response to a test item can be attributed to cultural differences.

Local independence (or conditional independence) can be investigated by examining the residual correlations of items within a scale.[4] A test is uni-dimensional when "the statistical dependence among items can be accounted for by a single latent trait".[22] Confirmatory factor analysis (CFA) should be applied to determine whether just one latent variable is defined on the basis of the correlation structure of the items within one scale, in other words whether the uni-dimensionality criterion is met.[5] As the items considered here cannot be assumed to be normally distributed, polychoric correlation matrices are used to assess uni-dimensionality as outlined by Samejima.[16]

In applying Confirmatory Factor Analysis (CFA) to assess uni-dimensionality there are two aspects to be considered: the estimated coefficients of each of the items in the structural equations defining the latent variable and global goodness of fit statistics. For the first aspect, the *t*-statistics corresponding to each of the loading coefficients must be statistically significantly different from zero. Three global goodness of fit statistics which could be considered are the Root Mean Square Error of Approximation (RMSEA) as suggested by Shevlin et al.[23]; the Tucker-Lewis or Non-Normed Fit Index (NNFI) recommended by Marsh et al.[24] as being appropriate for relatively smaller sample sizes, and the Comparative Fit Index (CFI) suggested by Bentler [25]. RMSEA values of less than .05 are good and greater than 0.10 poor while NNFI values greater than 0.90 are good and greater than 0.95 are excellent and CFI values greater than 0.93 are ideal.

## Evaluating model fit of the IRT model

Having met the required assumptions, the item parameters can be estimated. A relatively parsimonious model is desired without forfeiting too much information. In practice this entails comparing the fit of two models, one with a fixed and one with a variable slope or discrimination parameter, since a fixed slope parameter means a reduction in the number of estimated parameters. Chi-square statistics are applied to compare these nested models. Sample size and skewed distributions may hinder obtaining good estimates which can be remedied by setting prior parameter values to arrive at reasonable estimates.[26]

# References

1. Smith PA, Davis AC, Pronk M, Stephens D, Kramer SE, Thodi C, Anteunis LJ, Grandori F. Adult hearing Screening: What comes next? (Letter to the Editor) *Int J Audiol* 2011;50:610-612.
2. Reeve BB, Hays RD, Chang C-H, Perfetto EM. Applying item response theory to enhance health outcomes assessment. *Qual Life Res* 2007;16:1-3.
3. Fayers PM. Applying item response theory and computer adaptive testing: the challenges for health outcomes assessment. *Qual Life Res* 2007;16 Suppl. 1:187-194.
4. Hays RD, Liu H, Spritzer K, Cella D. Item response theory analyses of physical function items in the medical outcomes study. *Med Care* 2007;45(5 Suppl 1): S32-38.
5. Edelen M, Reeve B. Applying item response theory (IRT) modelling to questionnaire development, evaluation, and refinement. *Qual Life Res* 2007;16:5-18.
6. Takene Y, De Leeuw J. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika* 1987;52:393-408.
7. Knol DL, Berger MPF. Empirical comparison between factor analysis and multidimensional item response models. *Multivar Behavl Res* 1991;26:457-477.
8. Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I. Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models. *Health Technol Asses*s 2007;11:1-294.
9. Manchaiah VKS, Stephens D, Meredith R. The patient journey of adults with hearing impairment: the patients' views. *Clin Otolaryngol* 2011;36:227-234.
10. Demorest ME, Wark DJ, Erdman SA. Development of the screening test for hearing problems. *Am J Audiol* 2011;20:100-110.
11. Mokkink LB, Knol DL, van Nispen RMA, Kramer SE. Improving the quality and applicability of the Dutch scales of the communication profile for the hearing impaired using item response theory. *J Speech Lang Hear Res* 2010;53:556-571.
12. Hallam RS, Brooks DN. Development of the Hearing Attitudes in Rehabilitation Questionnaire (HARQ). *Br J Audiol* 1996;30:199-213.
13. Chenault MN, Anteunis LJC, Berger MPF. Evaluation of the Hearing Aid Rehabilitation Questionnaire in Dutch: examination of its psychometric properties and potential use as a screening instrument. *Audiology Research* 2013;*3,*e5 32-41.
14. Jöreskog KG. A general approach to confirmatory factor analysis. *Psychometrika* 1969;34:183-202.
15. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Mongraph Supplement*, 1969:No. 17.
16. Samejina F. The graded response model. In: Van der Linden, W.J., & Hambleton, R. eds. *Handbook of Modern Item Response Theory*. New York, NY Springer; 1997:85-100.
17. Jöreskog KG, Sörbom D. LISREL 8.7 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International, Inc. 2004.
18. Thissen D, Chen W-H, Bock RD. MULTILOG 7 for Windows: Multiple-category item analysis and test scoring using item response theory [Computer software]. Lincolnwood, IL: Scientific Software International, Inc. 2003.
19. Baker FB. *The basics of item response theory*. Portsmouth. NH: Heineman. (online: http://echo.edres.org:8080/irt/baker/chapter6.pdf. 1985.
20. Flannery WP, Reise SP, Widaman KF. An item response theory analysis of the general and academic scales of the Self-description Questionnaire II. *J Res Pers* 1995;29: 168-188.
21. Thomas ML. The value of item response theory in clinical assessment: A review. *Assessment* 2011;18:291-307.
22. Crocker L, Algina J. *Introduction to classical & modern test theory*, Harcourt Brace Jovanovich College Publishers, Orlando, Florida, U.S.A. 1986.
23. Shevlin M, Miles JNV, Lewis CA. Re-assessing the fit of the confirmatory factor analysis of the multidimensional students life satisfaction scale: comments on 'confirmatory factor analysis of the multidimensional Students' Life Satisfaction Scale'. *Pers Individ Dif* 2007;28:181-185.

24. Marsh HW, Bailla JR, McDonald RP. Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychol Bull* 1988;103:391-410.
25. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990;107:238-246.
26. Rupp A. Item response modeling with BILOG-MG and MULTILOG for Windows; *Int J Testing* 2003;3: 365–384.

3

# Chapter 4

An investigation of measurement equivalence in hearing response scales: Refinement of a questionnaire for use in hearing screening

MN Chenault
LJC Anteunis
B Kremer
MPF Berger

# Abstract

**Purpose**
Questionnaires used in hearing screening should be short and demonstrate measurement equivalence across groups defined by hearing impairment and hearing aid experience. The measurement equivalence of 2 scales addressing *functionality* (experienced hearing ability) and *social hearing* (social barriers due to hearing problems) was investigated.

**Methods**
Measurement equivalence was assessed using the differential item functioning and differential test framework within item response theory. Three comparisons were considered: (a) persons with normal hearing versus persons with hearing impairment either owning or not owning a hearing aid; (b) hearing aid users versus individuals who do not use hearing aids; (c) hearing aid users versus persons with hearing impairment who do not use a hearing aid. A protocol of differential item detection was applied consisting of ordinal regression and the log-likelihood ratio test to flag suspect items, followed by applying the log-likelihood ratio test using anchor items (items not suspect of differential item functioning).

**Results**
The 11-item *functionality* scale was reduced to 9 items, whereas the 10-item *social hearing* scale was reduced to 7 items.

**Conclusion**
Applying the differential item functioning framework resulted in shorter questionnaires displaying measurement equivalence relative to hearing impairment and hearing aid use without loss of reliability.

## Introduction

It is estimated that in the developed world, at least one in 10 adults and 30% of those aged 65 years and above have hearing problems to the extent that they would benefit from some sort of support.[1,2] There is also abundant evidence that hearing rehabilitation is an intervention that enhances the functionality and general quality of life of the individual with hearing impairment.[3,4] Nevertheless, hearing aid uptake among those with hearing impairment is low. In a population aged 55 years and older, 40% reported complaints about hearing and 16% had consulted their general practitioner about it, whereas only 8% had been fitted with a hearing aid.[5] Moreover, many persons who have been fitted with hearing aids do not use them. In a large survey conducted among an older Australian population, 24% of the hearing aid owners reported that they never used their aid.[6] In a follow-up on first-time hearing aid users, Vuorialho et al. reported that 36.8% used their aids only occasionally and 5.3% did not use them at all.[4]

There is a considerable gap in time from the onset of hearing impairment at a level warranting attention to the moment at which help is sought. Manchaiah et al.[7] identified seven stages in the patient journey toward hearing rehabilitation: (a) pre-awareness, (b) awareness, (c) movement, (d) diagnostics, (e) rehabilitation, (f) self-evaluation, and (g) resolution. Help-seeking occurs in the movement stage after the person with hearing impairment has ascertained his or her limitations and decides to take action. Prior to this stage, the person with hearing impairment may typically exhibit behaviour including coping strategies or denial. A key point in this process, marking the beginning of the rehabilitation process, is when the person with hearing impairment consults caregivers regarding available solutions. Southall et al.[8] conducted a qualitative study among successful hearing aid users. This study revealed that a successful use of assistant technology involved four stages: recognizing one's hearing difficulties, being aware that there are technology solutions, consultation regarding the uptake of such technology devices, and (behavioural) adaptation to use. The term hearing assistance technology applied here includes any device that improves functional hearing. Although hearing aid uptake is generally perceived as being equivalent to hearing rehabilitation, there are other forms of rehabilitation. In a systematic literature review of studies regarding interventions offered following hearing screening, Pronk et al.[9] determined that alternative interventions with communication programs should be considered as either an addition or a replacement to hearing aid provision. Hickson et al.[10] evaluated the effectiveness of the Active Communication Education program for older persons and demonstrated that it could be considered as an alternative or supplement to hearing aid fitting. The recognition of the needs and the range of possibilities for rehabilitation by both the person with hearing impairment and his or her associated caregivers is possibly crucial to the success of any form of hearing rehabilitation.

4

It has been shown that, although there is an average delay of 10 years from the moment when intervention would provide the best results, earlier intervention means greater gains in effective rehabilitation.[1] For this reason, more attention is being paid to the role that screening could play in targeting those individuals who might benefit from timely rehabilitation. Although hearing screening may result in doubling[1] or even tripling the number of hearing aid owners,[11] it will also result in a considerable increase in nonusers. Approximately 60% of persons who had been fitted with hearing aids after a population hearing screening were no longer using them at follow-up.[1,12] Joore et al.[13] reported that 11% of hearing aid users are dissatisfied with their hearing aid, which contributes to underuse and can result in non-use. It has been shown that focusing on this population of unsatisfied owners to increase hearing aid use is highly cost-effective.[4,14] To reduce the number of hearing aids lying in a drawer unused, a hearing screening should therefore not only include audiometric measurement but also assess latent factors such as experienced hearing problems and attitudes related to hearing loss, keeping in mind the objective of successful hearing rehabilitation.

Pacala and Yueh[15] argue that, given the high prevalence of hearing loss often accompanied by the individual's lack of recognition of it, hearing screening is highly feasible, given relatively low costs, ease, and safety for the individual. Hearing screening can combine questionnaire items with objective audiological outcomes. Yueh et al.[11] discussed the possibility of combining (a) an audioscope, a sound-emitting otoscope, to identify physiologic hearing loss and (b) questionnaire items addressing hearing limitations. Audiometry outcomes may provide harder evidence because responses to questionnaire items may reflect lack of awareness of hearing impairment. Nevertheless, in a large cohort aged 55–99 years, Sindhusake et al.[16] found that items from the shortened Hearing Handicap Inventory for the Elderly[17] had high sensitivity rates from 79% for mild hearing loss (25–40 dB), 86% for moderate hearing loss (40-60 dB), and 93% for marked hearing loss (>60 dB).

The foremost purpose of a questionnaire developed to be used as a screening instrument is to filter out persons who are experiencing hearing problems. A potential screening population will include persons with minimal to moderate hearing impairment, assuming that persons with severe hearing impairment are already using a hearing aid. However, a potential screening population may possibly include persons who have a hearing aid but no longer use it. Therefore, in any potential hearing screening population, there are basically two groups to consider: (a) persons with no or only minimal hearing loss not experiencing hearing disability who would gain little or no benefit from hearing aid rehabilitation and (b) persons with hearing loss who might benefit. In this latter group, two subgroups can be distinguished: (a) those with hearing impairment who have not been fitted with a hearing aid and (b) persons who have a hearing aid but are not using it effectively.

Because a screening population will include persons with varying degrees of hearing impairment as well as persons who have or have not been fitted with a hearing aid, any

instrument employed should be bias free with regard to the level of hearing impairment and/or the person having experienced hearing aid use. Here, the instrument employed was a questionnaire designed to measure experienced hearing difficulties. It is important that the instrument or tool exhibits measurement equivalence or, in other words, be bias free across groups defined by hearing impairment and hearing aid use.

The subject of bias or measurement equivalence in health outcome measures was addressed by McHorney and Fleishman,[18] who pointed out that lack of measurement equivalence over groups could result in either over- or under detection in a screening setting. Measurement equivalence can be assessed within what is known as the differential test framework, which was proposed by Raju et al.[19] This approach uses the theoretical framework of item response theory (IRT), whereby the respondents' responses to questionnaire items are modelled. Bias is indicated when probability response curves, and thus the estimated parameters describing these response curves, differ between groups.

The most widely used screening questionnaire, the HHIE-s, developed by Ventry and Weinstein[17] to be used in combination with pure-tone audiometry, consists of 10 questions regarding difficulties experienced in specific situations. Other questionnaires applied to hearing and hearing aid outcomes have primarily been designed to assess and facilitate the hearing rehabilitation process when the person who is experiencing hearing disability seeks help. Expected Consequences of Hearing Aid Ownership (ECHO), developed by Cox and Alexander[20], measures hearing aid expectations. Hallam and Brooks[21] developed the Hearing Aid Rehabilitation Questionnaire (HARQ) to assess attitudes related to the rehabilitation process on the premise that such attitudes influence the individual's perception of his or her handicap. The HARQ consists of 41 items dispersed over seven scales, with 21 items pertaining to hearing and 20 items pertaining to hearing aids. Given that these items address not only hearing difficulties but also attitudes toward hearing impairment, hearing aids, and whether pressure to be assessed has been experienced, the HARQ addresses a wider scope of relevant issues, rendering it potentially useful for hearing screening.

In a previous study, Chenault et al.[22] derived two scales from 20 HARQ hearing items for application in hearing screening; these two scales were named *functionality* (quantifying functional hearing) and *social hearing* (quantifying experienced social barriers from hearing impairment). In a related study, IRT methodology was applied to evaluate these two hearing scales.[23] It was shown that the items within each scale provide information for varying trait levels (degree of either functional hearing or experienced social barriers) and, furthermore, that these items discriminate between different trait levels, rendering the scales useful in a hearing screen setting.

Although responses to scale items and, thus, the trait values obtained through IRT modelling quantify experienced hearing, we would, on one hand, expect these responses to be related to objective audiometric outcomes and, on the other hand,

vary relative to whether the individual has experienced hearing aid use. Policy criteria for hearing rehabilitation interventions are mostly based on audiometric outcomes that correlate but are not equivalent to experienced hearing problems. At the same time, a health intervention such as audiological rehabilitation has an impact on self-assessment. Joore et al.[24] reported a response shift after individuals had undergone hearing aid fitting, with those individuals having become more aware of the difficulties that they had experienced before hearing aid fitting. It is therefore important to determine whether items in a scale and the scale as a whole demonstrate measurement equivalence across groups defined by hearing impairment that qualifies for hearing aid fitting (reimbursement is not the same in all countries) and/or hearing aid use.

The purpose of the present study was to assess whether these two HARQ hearing scales, *functionality* and *social hearing*, can be applied as a tool for screening in a population of persons with varying degrees of hearing ability and with or without experience in wearing a hearing aid. In other words, do these two scales demonstrate measurement equivalence with respect to level of hearing impairment and hearing aid use? Often, the terms measurement equivalence and bias free are used interchangeably and, in the context here, apply only to hearing impairment and hearing aid use. To supplement the theory given in the subsequent section, an introduction into some basic aspects of IRT has been provided in Appendix I. An introduction to the application of IRT to audiology-related scales can be found in the Appendix of Demorest et al.,[25] in which the authors applied IRT to the Communication Profile for the Hearing Impaired (CPHI) with dichotomous response categories (yes/no). Also, readers are referred to the Appendix of the previously published related article by Chenault et al.,[23] in which more details on IRT applied to the HARQ hearing items are given.

## Materials and methods

### Differential item/test framework

Within IRT methodology, there is a modelling framework for determining whether items and scales exhibit bias across groups.[26] It is important to differentiate between differential test functioning (DTF) and differential item functioning (DIF). DTF determines whether the scale as a whole demonstrates bias across groups, whereas DIF investigates whether individual scale items are biased across groups. It is possible that a scale contains items exhibiting DIF and the scale as a whole does not exhibit DTF. This happens when individual items with DIF cancel out each other's DIF effect.

There are two types of DIF: uniform and non-uniform. Uniform DIF occurs when the probability of a response to a questionnaire item is different for two groups, but the rate at which the response changes relative to the underlying trait is equal across these

two groups. Non-uniform DIF occurs when the rate of response change relative to underlying trait differs across groups. Inspection of the item characteristic curve (ICC), which describes the probability of a response relative to the underlying trait being assessed, is useful for understanding the nature of DIF when it is present. If the ICCs of two groups are parallel but have varying threshold values, we speak of uniform DIF, the difference between response curves being constant relative to the underlying trait. Non-uniform DIF is said to occur when the ICCs vary in shape, which occurs when the groups' ICCs have markedly different slopes or discrimination parameters and, therefore, are not parallel.

The group being investigated for bias, referred to as the focal group, is compared to a reference group, which could be the rest of the sample. As an example, let us consider as our focal group those persons who are classified as having normal hearing, and let us consider as our reference group persons with hearing impairment and responses to item s9 from the HARQ *social hearing* scale: "My hearing loss makes me feel isolated from other people." The question is whether the response curves for the focal group are different from those of the reference group. In Figure 4.1, the top graph presents the ICCs for the reference group of persons with hearing impairment (solid line) and the focal group of persons with normal hearing (dotted line). The left curve for each of the groups, which slopes downward from left to right and describes the probability of responding with "agree" to this item, coincide. The probability of a response of "partly agree" is depicted by the middle curves and "disagree" by the curves to the right. These two response curves for the group with normal hearing are slightly to the right of those for the group of individuals with hearing impairment. However, these differences are negligible, as confirmed by statistical testing; therefore, it can be concluded that the ICCs for the two groups do not differ and that DIF is not present for this item. Let us now consider the group of hearing aid users as the focal group and, as the reference group, those individuals who do not use a hearing aid. In the middle panel of Figure 4.1, response curves show that these two groups differ. In particular, the right curves corresponding to the probability of responding with "disagree" are different not only in location but also in rate of change. Here, we speak of non-uniform DIF. The reference group (no hearing aid use) disagrees more readily with this statement than does the focal group (hearing aid users), possibly reflecting the experience of hearing aid users. Now let us consider item s6: "I am sure that some people think I am stupid just because I have a hearing loss" and the comparison of persons with normal hearing (focal group) versus persons with hearing impairment (reference group). When comparing the ICCs, we see an example of uniform DIF. In particular, the middle response curve corresponding to "partly agree" is different for the group with normal hearing and is located clearly to the right of the curve for the group of individuals with hearing impairment. Because the three curves are otherwise parallel, we speak of uniform DIF. The rate of response is the same relative to the underlying trait, *social hearing*, which is depicted by the horizontal axis, with the group of individuals with hearing impairment

4

more likely to respond with either "partly agree" or "disagree" for the same trait levels than the normal hearing group.
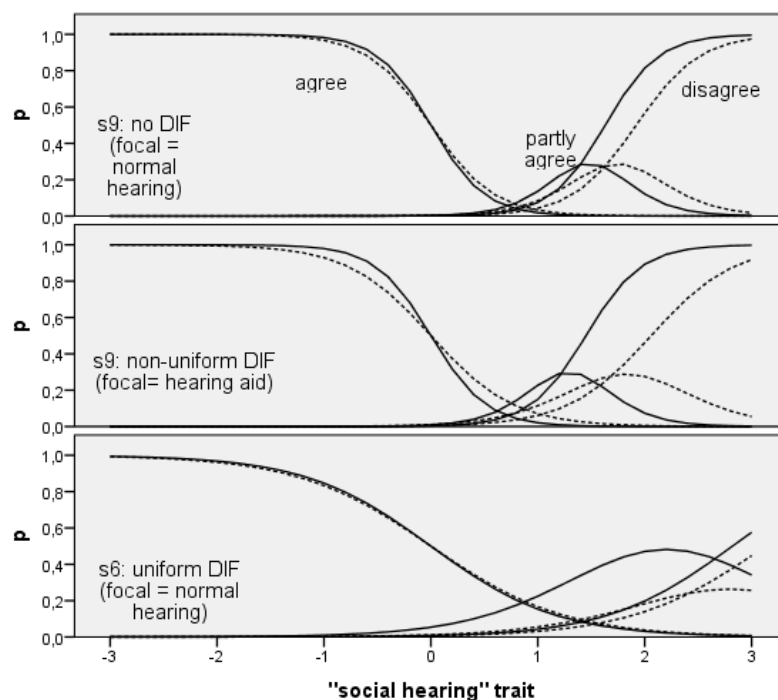
Figure 4.1      Item characteristic curves (ICCs) illustrating (a) absence of differential item functioning (DIF) pertaining to item s9, (b) non-uniform DIF pertaining to item s9, and (c) uniform DIF pertaining to s6.

In examining DIF, then, there are three basic elements to consider: (a) item responses, (b) group membership, and (c) (dis)ability level, as it is the effect of group membership on response relative to ability that determines whether there is DIF or not. Dorans and Holland[27] explained the importance of including a variable to quantify the underlying trait (disability). This underlying trait can be a trait score obtained through IRT or classical test theory (CTT). Crane et al.[28] advised the use of sum scores obtained from CTT.

## Methods for determining DIF

Various methods have been introduced to assess the presence of DIF. These methods can be divided basically into IRT- and non-IRT–based approaches. Teresi[29] conducted a review of methods to detect DIF in health applications, whereby the advantages and disadvantages of these approaches were outlined. She recommended the IRT approach called the log-likelihood ratio test (IRTLR) by Thissen et al.[30] As for non-IRT approaches,

Teresi[29] stated that the logistic regression (LR) approach introduced by Swaminathan and Rogers[31] for binary outcomes performs well and shows greater power for detecting DIF than other non-IRT methods. As reported by Lai et al.,[32] it is not only common but advisable to use more than one approach to assess DIF and that different approaches may complement each other. In the present study, the LR and IRTLR approaches were applied.

*Ordinal LR method*

The LR approach is appropriate when there are two possible outcomes. Zumbo[33] extended this approach for ordinal item responses. The ordinal logistic regression (OLR) method considers models with the observed ordinal item responses (y) as the dependent variable. Three models can be distinguished. In the first model, only the ordinal response is estimated with the trait score. In the second model, the group variable for which the presence of DIF is being investigated is added as independent variable. In the third model, the interaction between the trait scale score and the group variable is included, in addition to the trait scale score and the group variable. The corresponding coefficients can be interpreted as follows: $\beta_1$ is the estimate of the coefficient for trait score (T), $\beta_2$ is the estimation of the mean difference in response between groups (G) conditioned by $b_1$ on the trait score (T), and $b_3$ is the coefficient for the interaction term (T × G) between the group and trait score (T). A model fit difference between Models 2 and 3 indicates non-uniform DIF, and a model fit difference between Models 1 and 2 indicates uniform DIF:

Model 1:  $y = \beta_0 + \beta_1 T$

Model 2:  $y = \beta_0 + \beta_1 T + \beta_2 G$

Model 3:  $y = \beta_0 + \beta_1 T + \beta_2 G + \beta_3 (T \times G)$

There are two steps to this method. In the first step, Models 2 and 3 are compared, whereby the chi-square model fit test is applied with a Type I error of α=.05. If the change in model fit is not statistically significant, it can be assumed that non-uniform DIF is absent and only then is uniform DIF investigated, in Step 2, where Models 1 and 2 are compared. If the fit of these two models are not significantly different, uniform DIF is not assumed.

*IRTLR method*

The IRTLR method involves a comparison of two IRT models using the likelihood ratio test statistic. The model fits statistics obtained when parameters across groups that are constrained to be equal are compared to model parameters whereby the parameters for the item suspected of having DIF are estimated separately. There is DIF if the difference in fit between the two models is statistically significant. Analyses are

**4**

performed to determine which items appear to be DIF-free so that these items can be considered anchor items. These anchor items are then constrained to be equal across groups, allowing only the estimated parameters of the item in question to vary across groups. Two models are then compared: the model with the unconstrained potential DIF item and the model in which all the items are constrained to be equal across groups. A chi-square test of the difference between the fit of the two models tests whether that item has DIF. (See Thissen et al. for more details.)[30]

## Interpretation of DIF and consequences

Because evidence of DIF indicates a threat to measurement equivalence, its presence can be a reason for dropping an item from a scale. Although statistically significant differences in model fit provide evidence of DIF, the examination of graphs such as ICCs (see explanation regarding Figure 4.1) and information curves across groups is useful in understanding the nature of the DIF. The ICCs of the two groups will more or less coincide when there is no DIF. Uniform DIF is present when the ICCs do not coincide but are parallel, and non-uniform DIF is present when they are not parallel.

Although the ICC presents the probability of response relative to trait values, information curves plot the amount of information provided by an item per trait value. In the presence of DIF, the amount of information generated will differ between focal and reference groups, relative to trait values. At the same time, it is important to investigate whether the scale as a whole exhibits DTF. This is a separate analysis and basically involves a comparison of estimated trait values with and without suspected DIF items across groups.

## HARQ questionnaire

The *functionality* and *social hearing* scales consisted of 11 and 10 items, respectively, with all items measured on a 3-point ordinal response scale ("agree", "partly agree", and "disagree"); a higher score corresponded to fewer hearing problems. The graded response model (GRM) by Samejima[34,35] with a variable slope was specified to obtain the estimated IRT parameters for these two scales, which are presented in Appendix II.[23]

## Sample description

The sample consisted of responses from 212 Dutch persons, aged 55 years or older, who were recruited through a large general practice clinic in the southern region of the Netherlands. The definition of hearing impairment applied here is based on the criterion for hearing aid reimbursement in the Netherlands, which, at the time of data collection, was a pure-tone average best ear (PTABE) of 35 dB or more for the frequencies 1, 2, and 4 kHz. Our sample includes 85 persons who did not have hearing

impairment and were considered to have normal hearing, and 127 persons with hearing impairment according to this definition. Within the group of individuals with hearing impairment, 63 persons were hearing aid users who had been included because any potential hearing screen population will also include persons who have been fitted with a hearing aid. Our assumption was that these persons would capture the experience of those who had gone through the process of hearing aid fitting, which includes the realization or acceptance of their hearing impairment.

## Procedure

The procedure followed for the detection of DIF echoes that proposed by Teresi and Fleishman,[36] which consists of the identification of anchor items (items clearly not exhibiting DIF), investigation of items suspect of DIF, parameter estimation for a (reduced) set of items without DIF, and evaluation of the impact of DIF by comparing reduced scales with original scales.

Given the three groups thus defined, there were a number of comparisons to be considered to investigate the presence of item and scale bias. Because our objective was to obtain hearing scales to use in hearing screening that are free of bias relative to hearing impairment and hearing aid use, we selected three comparisons for this purpose. The hypothesis to be tested at each step was whether DIF occurred in a focal group when compared to a reference group. The three comparisons are as follows:

- Comparison I: Focal group consisted of persons with normal hearing (n=85), PTABE (dB), median=25, interquartile range (18,28) dB. Reference group (n=127) consisted of persons with hearing impairment, with or without a hearing aid.
- Comparison II: Focal group consisted of hearing aid users (n=63), PTABE (dB), median=51, interquartile range (42,62). Reference group (n=149) consisted of persons with either normal hearing or with hearing impairment but not using a hearing aid.
- Comparison III: Focal group consisted of hearing aid users (n=63). Reference group (n=64) consisted of persons with hearing impairment but not using a hearing aid, PTABE (dB), median=43, interquartile range (38,48).

In the first comparison, measurement equivalence was investigated by comparing the normal-hearing group (mean PTABE<35 dB over frequencies 1, 2, and 4 kHz) to the rest of the sample. The rationale for this comparison was that the items and their scales should not be biased relevant to audiometrically measured hearing ability because the object of the questionnaire is to quantify subjectively experienced hearing ability.

In the second comparison, the groups compared were those with and without a hearing aid. Within the group of persons who did not use a hearing aid, individuals ranged from those with negligible hearing ability to those with moderate hearing impairment

(M=44.7 dB). Literature shows that persons vary according to the amount of hearing complaints relative to objectively obtained audiometric outcomes[37,38].

In the third comparison, because of the substantial percentages of individuals who did not use a hearing aid reported in populations of persons with hearing impairment, only the individuals with hearing impairment were compared to hearing aid users.

*Step 1. Investigating non-uniform DIF.*
In the first step, the search for items without non-uniform DIF was conducted for each of the two HARQ hearing scales: *functionality* and *social hearing*. Both the IRTLR and the OLR methods using ordinal regression were applied to flag items that were suspect of non-uniform DIF. For both methods, a chi-square statistic greater than 3.86 corresponding to α=.05 was considered statistically significant. This investigation was conducted for each of the three comparisons. The items with no indication of non-uniform DIF were considered to be anchor items for a re-examination of non-uniform DIF using the IRTLR method. Any suspect item for which DIF was confirmed was dropped from its respective scale.

*Step 2. Investigating uniform DIF.*
Thereafter, the process was repeated to assess the presence of uniform DIF. In comparing Model 1 with Model 2, the relevant chi-square statistic was still 3.86 for the OLR method. For the IRTLR method, however, the chi-square statistic depends on the number of threshold parameters estimated. Because the responses were set on a 3-point scale, there were two threshold parameters estimated, resulting in 2 degrees of freedom, so that a chi-square statistic greater than 6.0 corresponds to α=.05. In the event that an item had zero counts for a response category in a group considered, thus estimating only one threshold parameter, the responses for the reference group were merged. In this case, a chi-square statistic exceeding 3.86 corresponds to α=.05.

An adjustment for multiple testing was not applied when determining the set of anchor items as recommended by Thissen.[37] Multiple comparisons may result in not removing DIF items from a scale, while the objective was to derive a set of pure anchor items without DIF.[28] A Bonferroni correction for multiple testing was applied when performing the IRTLR analysis of items suspect of DIF using the anchor items.

Statistical testing may provide direction, but it is the inspection of the items relative to their estimated parameters that provides insight into the nature of the lack of measurement equivalence. ICCs and information curves for the reference and focal groups were therefore compared to examine the nature of DIF for the items where lack of measurement equivalence was found. When DIF is not present, the ICCs will (almost) coincide, but when DIF is present, they will not coincide because of the differences in either estimated discrimination or threshold parameters. Because information curves display the amount of information an item provides per trait level, DIF is indicated

when the information curves vary in height relative to trait between groups, as is the case in non-uniform DIF or in a location along the trait scale when uniform DIF is present.

DTF, or the impact of excluding items with DIF, was investigated by comparing scale scores with and without items detected as having DIF for each of the groups. Cronbach's alphas (internal consistency) for the scales with and without items demonstrating DIF were compared because generally shorter scales will result in lower reliability coefficients.[38] Data analysis was performed with Multilog 7.3,[39] IRTPRO,[40] and IBM SPSS (Version 21.0; IBM, Armonk, NY).

## Results

### Analysis of *functionality* scale items

#### Non-uniform DIF analysis of functionality scale items

In Table 4.1, the results of the DIF analysis of the *functionality* scale are presented, with three sections: results for Comparison I (left), Comparison II (centre), and Comparison III (right). The upper portion of the table pertains to the investigation of non-uniform DIF. The chi-square test statistics in the first column of each comparison are from the IRTLR analysis; the rest of the items in the scale are used as anchor items to examine DIF for each item separately. The chi-square test statistics in the second column are the results from the OLR method. The chi-square statistics in the third column pertain only to the items being investigated further for DIF, and the anchor items (in bold) are those items in each comparison that are not suspect of exhibiting DIF according to either method.

The IRTLR method flagged item f1, "It sometimes depresses me when I cannot follow a conversation" for Comparisons II and III; item f4, "I find myself avoiding company because conversation is too much effort" for Comparison II; item f5, "In a conversational group, I keep quiet for fear of saying the wrong thing" for Comparison III; and item f7, "It really upsets me when I realize I've got the 'wrong end of the stick' in conversation" for Comparison I. The OLR method flagged item f3, "I think I've overcome any hearing difficulties I might have through my own efforts" for Comparisons I and II and item f1 for Comparison III. The items that were not flagged by either approach for the three comparisons and, therefore, could be considered to be anchor items were f2, f6, f8, f9, f10, and f11. These six anchor items were then used to further investigate the five items suspected of non-uniform DIF for each of the relevant comparisons separately.

The subsequent IRTLR analysis of the five *functionality* items suspect of non-uniform DIF using the six anchor items indicated that only item f7 demonstrated non-uniform

DIF, and then only for Comparison I when the focal group was the group of individuals without hearing impairment.

*Uniform DIF analysis of functionality scale (omitting item f7)*

In the lower part of Table 4.1, the results of the uniform DIF analysis are presented. Both the IRTLR and OLR methods detected uniform DIF for the reverse-coded item f11, "Hearing is not a major concern for me at the moment" for Comparisons I and II, and the OLR method detected it also for item f4 for Comparison II. When the remaining eight items were taken as anchor items and the IRTLR method was applied, uniform DIF was confirmed for item f11 for Comparison I.

Table 4.1    DIF analysis (chi-square statistics) for IRTLR, OLR methods & IRTLR with anchor items (in bold) for *functionality* scale for Comparison I (focus: non-hearing impaired), Comparison II (focus: hearing aid), Comparison III (focus: hearing aid, reference group: hearing impaired, no aid).

| Item | Comparison I | | | Comparison II | | | Comparison III | | |
|---|---|---|---|---|---|---|---|---|---|
| | IRTLR | OLR | IRTLR | IRTLR | OLR | IRTLR | IRTLR | OLR | IRTLR |
| | all | all | anchor | all | all | anchor | all | all | anchor |
| non-uniform DIF | | | | | | | | | |
| f1 | 0.22 | 1.17 | | 5.10* | 2.25 | 2.33 | 8.07* | 5.38* | 3.06 |
| **f2** | 0.40 | 0.64 | | 1.98 | 0.00 | | 0.86 | 0.100 | |
| f3 | 2.87 | 16.49* | 0.84 | 0.04 | 4.63* | 0.03 | 0.60 | 0.55 | |
| f4 | 0.19 | 0.76 | | 3.94* | 0.04 | 2.60 | 1.68 | 0.83 | |
| f5 | 0.25 | 0.27 | | 2.33 | 1.12 | | 4.85* | 3.57 | 2.41 |
| **f6** | 0.04 | 0.01 | | 1.52 | 0.70 | | 1.98 | 1.22 | |
| f7 | 6.87* | 0.00 | 9.69[+] | 0.00 | 0.01 | | 0.03 | 0.75 | |
| **f8** | 0.04 | 0.01 | | 0.74 | 0.28 | | 0.38 | 0.28 | |
| **f9** | 0.06 | 0.03 | | 0.06 | 1.95 | | 0.23 | 2.38 | |
| **f10** | 0.17 | 1.65 | | 0.67 | 0.05 | | 1.84 | 0.77 | |
| **f11** | 3.10 | 1.00 | | 1.70 | 1.96 | | 0.17 | 0.63 | |
| uniform DIF | | | | | | | | | |
| **f1** | 2.00 | 1.46 | | 0.70 | 0.02 | | 0.52 | 0.01 | |
| **f2** | 1.10 | 0.17 | | 2.95 | 2.34 | | 1.43 | 2.00 | |
| **f3** | 2.01 | 0.40 | | 1.78 | 0.16 | | 4.62 | 2.04 | |
| f4 | 5.31 | 0.76 | | 5.32 | 4.06* | 5.12 | 2.90 | 2.53 | |
| **f5** | 0.39 | 0.27 | | 1.59 | 1.10 | | 3.35 | 2.99 | |
| **f6** | 0.72 | 1.00 | | 0.24 | 0.03 | | 0.11 | 040 | |
| **f8** | 0.25 | 0.85 | | 1.01 | 2.13 | | 1.41 | 1.33 | |
| **f9** | 2.10 | 3.42 | | 0.36 | 1.31 | | 3.93 | 0.07 | |
| **f10** | 0.68 | 0.26 | | 0.81 | 0.14 | | 0.77 | 0.01 | |
| f11 | 8.27* | 7.13* | 7.08[+] | 6.23* | 5.02* | 3.99 | 3.46 | 0.37 | |

* denotes a p-value < .05; [+] denotes a p-value < .05 after Bonferroni correction

*ICC and information curves of DIF items of functionality scale*

The ICC and information curves of items for the comparisons for which DIF was substantiated are presented in Figure 4.2. For item f7, the presence of non-uniform DIF for Comparison I is evident in that the ICCs are clearly different in shape.
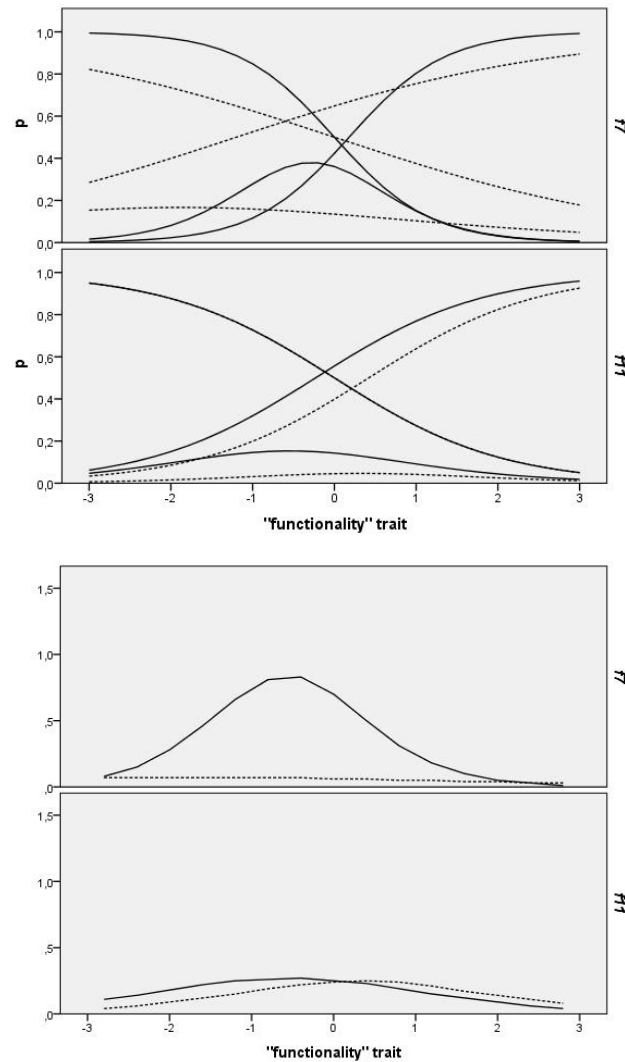
Figure 4.2    Panel a: Item characteristic curves (ICCs) for items f7 and f11 for the reference group versus the focal group pertaining to Comparison I. Panel b: Information curves for items f7 and f11 for the reference group versus the focal group pertaining to Comparison I.

The flatness in the ICCs for the group of individuals without hearing impairment reflects the lack of discrimination ability of this item in this group. Uniform DIF is visually present for item f11 in the shift to the right for the right curve, indicating that, for the group of individuals without hearing impairment, the crossover from the middle to the highest response category occurs at higher trait values. In Figure 4.2b, the information curve for item f7 for the focal group in Comparison I is very flat and close to zero,

indicating that this item provides almost no information in the group without hearing impairment due to the item's lack of discriminative ability in this group. The information curve of item f11 for the group of individuals without hearing impairment is shifted to the right, indicating that relatively more information is being provided at relatively higher trait values.

## Analysis of *social hearing* scale items

### *Non-uniform DIF analysis of social hearing scale items using anchor items*

The results for the *social hearing* scale are presented in Table 4.2. The IRTLR method detected item s4, "When you have hearing difficulties, other people ignore you," for Comparison I; item s8, "Some people avoid me because of my hearing ability," for all three comparisons; and item s9, "My hearing loss makes me feel isolated from other people," for Comparisons II and III. The OLR method did not detect any one item as potentially exhibiting non-uniform DIF for any of the comparisons. Items s1, s2, s3, s5, s6, s7, and s10 were then considered to be anchor items to further investigate non-uniform DIF.

Table 4.2    DIF analysis (chi-square statistics) for IRTLR, OLR methods & IRTLR with anchor items (in bold) for *social hearing* scale for Comparison I (focus: non-hearing impaired), Comparison II (focus: hearing aid), Comparison III (focus: hearing aid, reference group: hearing impaired, no aid).

| Item | Comparison I | | | Comparison II | | | Comparison III | | |
|---|---|---|---|---|---|---|---|---|---|
| | IRTLR all | OLR all | IRTLR anchor | IRTLR all | OLR all | IRTLR Anchor | IRTLR all | OLR all | IRTLR anchor |
| non-uniform DIF | | | | | | | | | |
| **s1** | 0.14 | 0.01 | | 0.14 | 0.04 | | 0.01 | 0.10 | |
| **s2** | 0.78 | 0.02 | | 0.67 | 2.71 | | 2.31 | 2.49 | |
| **s3** | 0.13 | 1.20 | | 0.81 | 0.21 | | 0.01 | 0.44 | |
| s4 | 6.49* | 0.40 | 9.72[+] | 0.01 | 0.36 | | 0.34 | 0.95 | |
| **s5** | 0.00 | 0.50 | | 1.16 | 0.03 | | 0.70 | 0.19 | |
| **s6** | 0.40 | 1.37 | | 0.02 | 0.17 | | 1.50 | 1.32 | |
| **s7** | 2.87 | 2.72 | | 3.23 | 2.32 | | 1.24 | 0.69 | |
| s8 | 6.73* | 0.34 | 0.31 | 4.82* | 1.56 | 4.31* | 5.17* | 0.11 | 0.31 |
| sf9 | 3.57 | 0.20 | | 4.96* | 0.50 | 3.39 | 4.66* | 1.49 | 0.21 |
| **s10** | 0.14 | 0.99 | | 0.08 | 0.54 | | 0.52 | 0.62 | |
| uniform DIF | | | | | | | | | |
| **s1** | 0.75 | 0.01 | | 1.12 | 0.01 | | 0.69 | 0.25 | |
| **s2** | 0.66 | 3.69 | | 1.31 | 0.51 | | 1.46 | 0.43 | |
| s3 | 3.71 | 19.85* | 2.75 | 4.52 | 12.49* | 4.79 | 1.96 | 2.18 | |
| **s5** | 0.49 | 0.27 | | 1.19 | 0.23 | | 1.27 | 0.36 | |
| s6 | 7.97* | 2.99 | 7.81[+] | 1.24 | 2.13 | | 0.11 | 0.24 | |
| **s7** | 0.91 | 0.49 | | 1.17 | 1.66 | | 0.92 | 0.38 | |
| s8 | 0.20 | 0.01 | | 3.01 | 4.38* | 3.17 | 5.32[1]* | 7.52* | 5.71[+] |
| **s9** | 3.07 | 0.45 | | 4.34 | 0.91 | | 2.43 | 3.55 | |
| **s10** | 1.78 | 0.16 | | 1.16 | 0.34 | | 0.33 | 0.63 | |

\* denotes a *p*-value <.05, [1]=1 df, [+] denotes a *p*-value <.05 after Bonferroni correction.

The subsequent DIF analysis, using these seven *social hearing* scale items as anchor items, confirmed the presence of non-uniform DIF for item s4 for Comparison I and item s8 for Comparison II; however, after a Bonferroni adjustment for multiple comparisons, non-uniform DIF was confirmed only for item s4.

*Uniform DIF analysis of social hearing scale (omitting s4)*

Taking the remaining nine *social hearing* items, the IRTLR method flagged item s6, "I am sure that some people think I am stupid just because I have a hearing loss," for Comparison I and item s8 for Comparison III. The OLR method indicated that (a) item s3, "I find myself avoiding company because conversation is too much effort," demonstrated uniform DIF for Comparisons I and II and that item s8 for Comparisons II and III. This left items s1, s2, s5, s7, s9, and s10 to be considered as anchor items. The subsequent IRTLR analysis using these six anchor items confirmed the presence of uniform DIF for item s6 for Comparison I and the presence of uniform DIF for item s8 for Comparison III.

*ICC and information curves of DIF items of social hearing scale*

The ICCs (see Figure 4.3a) for item s4 show the lack of discriminatory ability that this item has in the group of individuals without hearing impairment. However, this item had only an estimated discrimination parameter of 0.64 for the entire sample. The estimated discrimination parameter for the group of individuals without hearing impairment was just below zero in value in the DIF analysis, indicating that this item does not discriminate experienced social hearing problems at all within this group. Furthermore, the examination of the information curves shows that this item makes no contribution to the scale in this group. For items s6 and s8, both the ICCs shifted to the right for the focal group, indicating that the crossover from "partly agree" to "disagree" occurs at relatively higher trait values. The shift to the right in the information curves for the focal groups means relatively more information for higher trait values in the focal group (see Figure 4.3b).

## DTF

In Figure 4.4, standardized scale scores of the original and reduced (after removal of DIF items) scales are compared relative to hearing group. When comparing the range of scores for the three groups, it can be noted that the range of scores of the three groups generally overlap for both scales but that persons in the group of individuals without hearing impairment tend to have higher trait values. The disparity between the scales with and without the items with DIF is fairly uniform for all three groups for the *functionality* scale. It can also be seen that the disparity between the original and reduced *social hearing* scale scores is largest at the lower and upper end of the trait values. For the group of hearing aid users and the group of individuals with hearing

impairment, this occurs at the lower end of the trait scale, whereas for the group of individuals without hearing impairment, it occurs at the higher end of the trait scale.
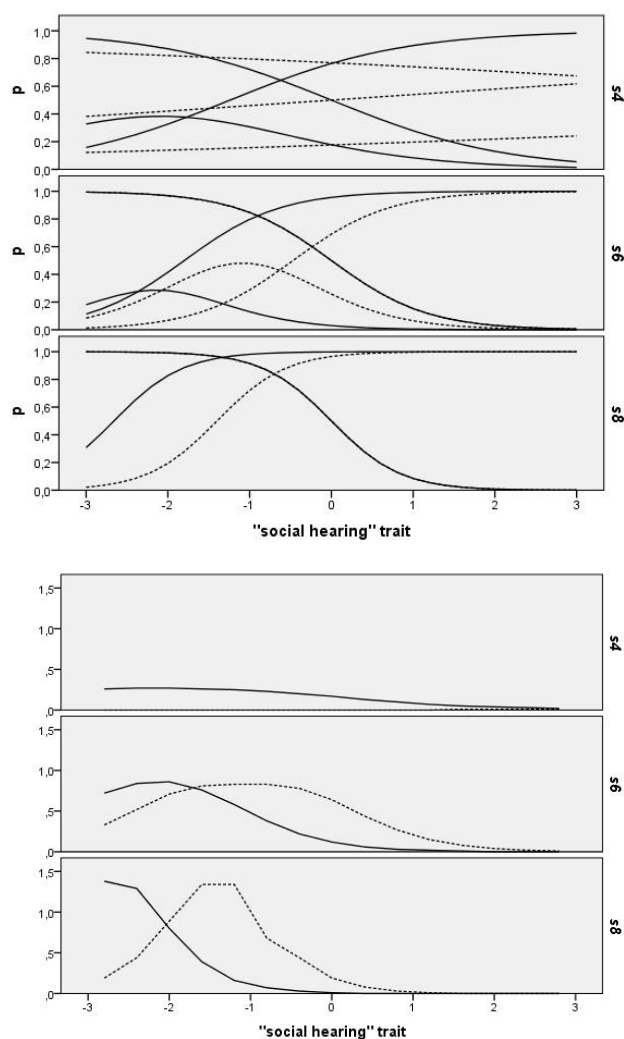


Figure 4.3    Panel a, top and middle graphs: Item characteristic curves (ICCs) for items s4 and s6 for the reference group versus the focal group pertaining to Comparison I. Panel a, bottom graph: ICCs for item s8 for the reference group versus the focal group pertaining to Comparison III. Panel b, top and middle graphs: Information curves for items s4 and s6 for the reference group versus the focal group pertaining to Comparison I. Panel b, bottom graph: Information curves for item s8 for the reference group versus the focal group pertaining to Comparison III.
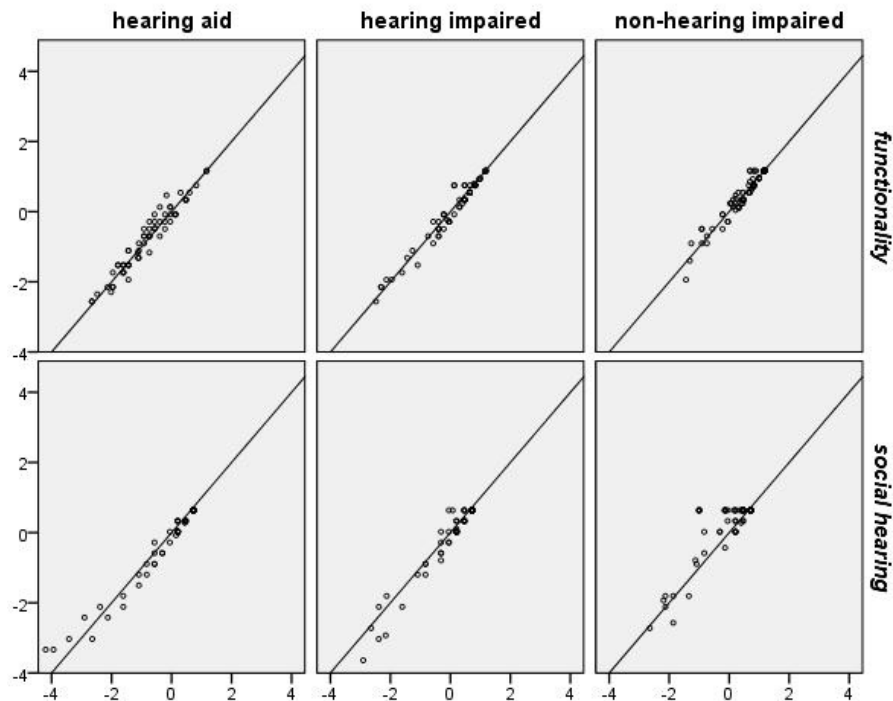
Figure 4.4    Comparison of standardized scale scores. (a) *Functionality* scale (top 3 panels): Nine items versus 11 items per group. (b) Social hearing scale (bottom 3 panels): Seven items versus 10 items per group.

*Final examination of scales after DIF analysis*

The final *functionality* scale consists of nine items namely, f1, f2, f3, f4, f5, f6, f8, f9, and f10; the *social hearing* scale consists of seven items namely, s1, s2, s3, s5, s7, s9, and s10. (For a full description of both sets of scale items, see Appendix II.) The deletion of items detected with DIF resulted in negligible decreases in the Cronbach's alpha for each scale. Reliability was .85 for the 11-item *functionality* scale and dropped to just .83 for a scale without item f7 and to .82 without item f11. For the 10-item *social hearing* scale, reliability was .85 and only dropped to .84 for the nine-item scale (without item s4) and to .83 for the seven item scale (without items s6 and s8).

## Discussion

In this study, IRT/DIF methodology was applied as a tool to refine two hearing scales. The goal was to derive scales with measurement equivalence to assess hearing in a

4

screen setting, with evidence that the scales would not be biased relative to hearing impairment and hearing aid use. The applied DIF detection procedure resulted in the 11-item *functionality* scale being reduced to nine items and the 10-item *social hearing* scale reduced to seven items. It can be assumed that these reduced scales exhibit measurement equivalence independent of participants having experienced hearing aid use and degree of hearing impairment.

For the *functionality* scale, non-uniform DIF was confirmed for item f7 and uniform DIF was confirmed for item f11 when the focal group was individuals without hearing impairment. The content of item f7, "It really upsets me when I realize I've got the 'wrong end of the stick' in conversation" which reflects having trouble understanding the context of a conversation may be attributable to non-audiological factors evoking "agree" responses from persons not otherwise experiencing hearing problems. For the reverse-coded item f11, "Hearing is not a major concern for me at the moment," inspection of the ICCs indicates that the second threshold is shifted to the right for the normal hearing group, indicating that persons with normal hearing are less likely to consider their hearing a major concern.

When examining the *social hearing* scale for bias in the normal-hearing group, non-uniform DIF was detected for item s4, "When you have hearing difficulties other people ignore you," and uniform DIF was detected for item s6, "I am sure that some people think I am stupid just because I have a hearing loss." The content of these two items is interpreted differently, depending on whether the respondent has had these experiences or not. For persons with normal hearing, it may address having had the experience of ignoring someone else with hearing problems or, possibly, considering that someone who does not hear well is "stupid." For the third comparison, in which hearing aid users were compared to persons with hearing impairment without a hearing aid, item s8, "Some people avoid me because of my hearing ability," displayed uniform DIF. For this comparison, only two types of responses were given, "partly agree" or "disagree", the first of which came from the group of hearing aid users. The ICCs and information curves show a shift to the right for the hearing aid users: In other words, at relatively higher levels of social hearing, hearing aid users cross over to a "disagree" response, which may reflect the experience of hearing aid use and the social barriers that may result from wearing a hearing aid.

It is interesting to point out that DIF was not substantiated for any of the items in the second comparison, in which the hearing aid users were compared to the rest of the sample. Inclusion of the third comparison occurred because of the ongoing discussion regarding the lack of hearing aid uptake in the population of persons with hearing impairment. This choice appears to be warranted by the fact that DIF was found for one item for this comparison. At the same time, it is interesting to note that all other DIF occurred in the first comparison, which compares responses from the group of persons without hearing impairment to the responses from the rest of the groups. This is important in light of the fact that the scales are intended for use in hearing screening in

which the primary objective is to single out those persons who would benefit from some form of hearing rehabilitation based on their experienced hearing problems. It is apparent, upon examination of the content of each of these four items exhibiting DIF, that they would either appear to be ambivalent or non-applicable for persons who do not experience hearing problems. The usefulness of DIF methodology to filter out items that are ambiguous across groups is evident. Since the primary objective of a hearing screening is to screen out those persons who have experienced problems, DIF detection has shown that these items are not as useful as the other items in their respective scales.

The DTF analysis compared the sum scores of the original scales and the scales reduced by eliminating items with confirmed DIF relative to hearing group. The reduced *functionality* scales indicated that the greatest disparities, although almost negligible, occurred for higher trait values in the groups of individuals with hearing impairment. The differences were greater for the *social hearing* scales, with the greatest differences occurring in the group of hearing aid users. These differences were not consistently positive or negative, which could be attributable to the elimination of items s4 and s8 due to DIF in different focal groups. The largest disparities were observed at the upper end of the trait values in the group of individuals without hearing impairment, which suggests the lack of applicability or appropriateness of the items removed for this group.

In the end, we are interested in the impact of such items on estimated trait values because it is these trait values that will play an important role in a screening setting. The scales are used to assess experienced hearing by quantifying the traits considered: functionality and social hearing. Low values indicate experienced hearing problems and, thus, potential benefit from hearing rehabilitation. The final decision as to whether an item should be omitted from a scale should take into consideration advice given by Hambleton,[41] who stressed the importance of examining the impact rather than relying solely on statistical tests. In comparing the scale scores with and without DIF items, we observed greater differences toward the outer bounds of the obtained scores, which appear to reflect the S-form of the response functions. However, greater differences were observed for higher trait values where relatively fewer hearing problems are experienced. In the end, scales with a smaller number of items but without DIF were obtained without concessions with regard to reliability (internal consistency).

Hearing impairment, which is determined with audiometry, is at the basis of policy guidelines for hearing aid reimbursement. However, given the fact that individuals vary in how they cope and adjust to hearing impairment, many researchers in the field of audiology have recognized the need for self-assessment scales to quantify the individual's perception of his or her hearing problems.[17,21,25] In line with this, Danermark et al.[42] discussed the need for instruments targeting disability that conform to the framework put forth by the World Health Organization in its International Classification of Functioning, Disability, and Health. Here we have adhered to the

premise that experienced hearing disability and hearing handicap, quantified by our *functionality* and *social hearing* scales rather than by audiometric outcomes, should be used to filter out persons who could benefit from rehabilitation. Furthermore, we adhere to the notion that such a screening instrument be free of bias relative to hearing impairment and hearing aids because any potential screening population will consist of persons with a wide range of hearing abilities and some who have experienced hearing aid use. DIF methodology has been applied here to evaluate items regarding their measurement equivalence across hearing groups. The results show that each scale could be reduced to a set of items free of bias relative to hearing impairment and hearing aid use, providing evidence that these scales are suitable for further application in a hearing screen setting. Reduced, bias-free scales are desirable not only from the perspective of measurement but also because it is more feasible to use shorter, more practical scales, and here it was shown that the reliability of the scales was not compromised.

## Appendix I

### Supplemental information regarding the application of Item Response Theory (IRT) to the scales *functionality* and *social hearing*

There are two types of parameters which are estimated in IRT models: discrimination parameters (a) and threshold parameters (b). Higher discrimination parameters indicate better discriminative ability. Higher thresholds indicate that higher trait values are required for the transition from one response category into the adjacent higher response category. Here the responses to the items of the *functionality* and *social hearing* scales are ordered so that the trait is on a scale from low to high with higher values indicating relatively less impediment in hearing functioning and hearing in social situations.

The items considered here are on a 3-point ordinal scale and therefore an IRT model proposed by Samejima called the Graded Response Model (GRM) is appropriate, and the slopes will be allowed to vary per item.[34,35]

As an example, consider a HARQ item such as: "When several people are chatting, it bothers me that I often lose the thread of the conversation". With IRT modeling we estimate the trait values of hearing functionality which mark the transition from the responses "agree" to "partly agree" and from "partly agree" to "disagree". IRT estimation furthermore determines how sharp that transition is, in other words how distinct adjacent response categories are from each other. Higher values for the discrimination or slope parameter indicate more discriminatory power, quantifying the change in response relative to trait level. In a practical setting this means that items with higher discriminatory parameters can clearly pinpoint those persons who have relatively more problems. In a study applying the graded response model, Flannery et al. stipulated that values $a<0.8$ are unacceptable, $0.8<a<1.0$ are marginally acceptable and $a>1.0$ are acceptable.[43]

Given that the items considered here are on a 3-point scale: "agree", "partly agree", and "disagree", there are two threshold parameters ($b_1$ and $b_2$) to be estimated whereby $b_1$, marks the 50% chance of crossing from the lowest response category "agree" a higher category while the second threshold, $b_2$, marks the 50% chance transition from the joined categories of "agree" and "partly agree" to "disagree".

The Item Characteristic Curve (ICC) of each item provides a graphical representation of the item's discriminative ability and thresholds. The horizontal scale is the trait scale which is often symbolized by the Greek letter theta. The ICC curves vary in their placement on the horizontal or trait axis which determines the item specific estimated threshold values.

The item response model, as specified by the discrimination and threshold parameters, results in an item characteristic function which is visualized in an item characteristic curve (ICC) which describes the probability of response relative to the trait. This relation can be transformed into an item information curve. The information curve

plots the amount of information provided by each item relative to trait values. Information curves can range in form from being very peaked for a narrow range to being rather flat over a wide range of theta (trait) values. For diagnostic purposes a peaked information curve is more desirable. When the goal is to classify patients on a wider continuum, a peaked information curve may be less useful than one where the information is relatively constant over a wide range.[44] A thorough and clear explanation of the information curve can be found online (http://echo.edres.org.8080/irt/baker/).[45]

4

## Appendix II

| | Item | a | b1 | b2 |
|---|---|---|---|---|
| *functionality* | f1 It sometimes depresses me when I cannot follow a conversation | 1.86 | -0.39 | 0.17 |
| | f2 I dread meeting new people since becoming hearing impaired | 2.03 | -1.73 | -1.27 |
| | f3 I think I've overcome any hearing difficulties I might have through my own efforts | 0.78 | -0.37 | 0.52 |
| | f4 I find myself avoiding company because conversation is too much effort | 2.86 | -1.20 | -0.87 |
| | f5 In a conversational group I keep quiet for fear of saying the wrong thing | 2.47 | -1.05 | -0.59 |
| | f6 When several people are chatting, it bothers me that I often lose the thread of the conversation | 2.03 | -0.38 | 0.14 |
| | f7 It really upsets me when I realize I've got the "wrong end of the stick" in conversation | 1.47 | -1.13 | -0.67 |
| | f8 I can hear well enough when I really concentrate *(reversed)* | 2.01 | -1.88 | -1.00 |
| | f9 By and large I am able to hear without difficulty *(reversed)* | 3.39 | -1.12 | -0.62 |
| | f10 My hearing problems are really quite minor (*reversed*) | 1.46 | -0.31 | 0.27 |
| | f11 Difficulty in hearing is not of major concern to me at the moment (*reversed)* | 1.33 | -1.17 | -0.78 |
| *social hearing* | s1 I dread meeting new people since becoming hearing impaired | 2.37 | -1.57 | -1.12 |
| | s2 My poor hearing sometimes makes me feel really inadequate | 1.56 | -1.63 | -0.97 |
| | s3 I find myself avoiding company because conversation is too much effort | 2.73 | -1.21 | -0.85 |
| | s4 When you have hearing difficulties, other people ignore you | 0.64 | -4.14 | -1.65 |
| | s5 As I see it, I am less of a person because of my hearing difficulty. | 2.68 | -2.09 | -1.72 |
| | s6 I am sure that some people think I am stupid just because I have a hearing loss | 1.48 | -2.80 | -1.85 |
| | s7 I get the feeling that other people find it a strain to talk to me | 1.54 | -2.20 | -1.34 |
| | s8 Some people avoid me because of my hearing ability | 2.28 | -2.50 | -1.93 |
| | s9 My hearing loss makes me feel isolated from other people | 3.33 | -1.35 | -1.02 |
| | s10 I have to admit that deep down I feel restricted by my hearing loss | 2.43 | -1.07 | -0.65 |

*functionality* (*fi*) *and social hearing* scale (*si*) items and estimated IRT parameters (a: slope or discrimination, b(i), threshold parameters)

# References

1.  Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I. Acceptability, benefits and costs of early screening for hearing disability: a study of potential screening tests and models. *Health Technol Assess* 2007;11:1-294.
2.  Ries PW. Prevalence and characteristics of persons with hearing trouble: United States, 1990-91. Vital Health Stat 10 1994;188:1-75.
3.  Joore MA, Brunenberg DE, Chenault MN, Anteunis LCA. Societal effects of hearing aid fitting in the moderately hearing impaired. *Intl J Audiol* 2003;42:152-160.
4.  Vuorialho A, Karinen P, Sorri M. Effect of hearing aids on hearing disability And quality of life in the elderly. *Intl J Audiol* 2006;45:400-405.
5.  Duijvestijn JA, Anteunis LJC, Hendriks JJT, Manni JJ. Definition of hearing impairment and its effect on prevalence figures. A survey among senior citizens. *Acta Oto-Laryngologica* 1999;119:420-423.
6.  Hartley D, Rochtchina E, Newall P, Golding M, Mitchell P. Use of hearing aids and assistive listening devices in an older Australian population. *J Am Acad Audiol* 2010;21,642-653.
7.  Manchaiah VKS, Stephens D, Meredith R. The patient journey of adults with hearing impairment: The patients' views. *Clin Otolaryngol* 2011;36:227-234.
8.  Southall K, Gagne JP, Leroux T. Factors that influence the use of assistance technologies by older adults who have a hearing loss. *Intl J Audiol* 2006;45:252-290.
9.  Pronk M, Kramer SE, Davis AC, Stephens D, Smith PA, Thodi C, Grandori F. Interventions following hearing screening in adults: A systematic descriptive review. *Intl J Audiol* 2011;50:594-609.
10. Hickson L, Worrall L, Scarinci N. A randomized controlled trial evaluating the Active Communication Education Program for older people with hearing impairment. *Ear Hear* 2007;28:212-227.
11. Yueh B, Shapiro N, MacLean CM, Shekelle PG. Screening and management of adult hearing loss in primary care. *JAMA* 2003;289:1976-1985.
12. Gusselkoo J, de Bont L, von Faber M, Eekhof J, de Laat J, Hulshof J, Westendorp R. Auditory rehabilitation of older people from the general population; the Leiden 85-plus study. *Br J Gen Pract* 2003;53:536-540.
13. Joore MA, Van der Stel H, Peters HJM, Boas GM, Anteunis LJC. The Cost-effectiveness of hearing-aid fitting in the Netherlands. A*rch Otolaryngol Head Neck Surg* 2003;129:297-304.
14. Boas G, Van der Stel H, Peters H, Joore M, Anteunis L. Dynamic modeling in medical technology assessment; fitting hearing aids in the Netherlands. *International Journal of Technology Assessment in Health Care* 2001;17:618-625.
15. Pacala JT, Yueh B. Hearing deficits in the older patient. "I didn't notice anything". *JAMA* 2012;307: 1185-1194.
16. Sindhusake D, Mitchell P, Smith W, Golding M, Newall P, Hartley D, Rubin G. Validation of self-reported hearing loss. The Blue Mountains hearing study. *Int J Epidemiol* 2001;30:1371-1378.
17. Ventry IM, Weinstein BE. Identification of elderly people with hearing problems. *ASHA* 1983;28:37-42.
18. McHorney CA, Fleishman JA. Assessing and understanding measurement equivalence in health outcome measures. Issues for further quantitative and qualitative inquiry. *Med Care* 2006;44(11 Suppl 3):S205-10.
19. Raju NS, van der Linden WJ, Fleer PF. IRT-based internal measure of differential functioning of items and tests. *Applied Psychological Measurement* 1995;19:353-368.
20. Cox RM, Alexander GC. Expectations about hearing aids and their relationship to fitting outcome. *J Am Acad Audiol* 2000;*11:*368-382.
21. Hallam RS, Brooks DN. Development of the Hearing Attitudes in Rehabilitation Questionnaire (HARQ). *Br J Audiol* 1996;30:199-213.
22. Chenault MN, Anteunis LJC, Berger MPF. Evaluation of the Hearing Aid Rehabilitation Questionnaire in Dutch: examination of its psychometric properties and potential use as a screening instrument. *Audiol Res* 2013;*3,*e5:32-41.
23. Chenault MN, Berger MPF, Kremer B, Anteunis LJC Quantification of experienced hearing problems with item response theory. *Am J Audiol* 2013;22: 252-262.

24. Joore MA, Potjewijd J, Timmerman AA. Response shift in the measurement of quality of life in hearing impaired adults after hearing aid fitting. *Qual Life Re* 2002;11:299-307.

25. Demorest ME, Wark DJ, Erdman SA. Development of the screening test for hearing problems. *Am J Audiol* 2011;20:100-110.

26. Shealy RT, Stout WF. An item response theory model for test bias and differential test functioning. In P.W. Holland & H. Wainer (Eds.) Differential item functioning (pp 67-113). Hillsdale NJ: Lawrence Erlbaum, Inc. 1993.

27. Dorans NJ, Holland PW. DIF detection and description: Mantel-Haenzel and standardization. In P.W. Holland & H. Wainer (Eds). Differential item functioning (pp35-66). Lawrence Erbaum Associates: Hillsdale, NJ. 1993.

28. Crane PK, van Belle G, Larson EB. Test bias in a cognitive test. Differential item functioning in the CASI. *Stat Med* 2004; 23:241-256.

29. Teresi J. Different approaches to differential item functioning in health applications. Advantages, disadvantages and some neglected topics. *Med Care* 2006;*44*:152-170.

30. Thissen D, Steinberg L, Wainer H. Detection of differential item function using the parameters of item response models. In P.W. Holland & H. Wainer (Eds.) Differential item functioning (pp 67-113). Hillsdale NJ: Lawrence Erlbaum, Inc. 1993.

31. Swaminathan H, Rogers HJ. Detecting differential item functioning using logistic regression procedures. *Int J Educ Res* 1990;13:127-143.

32. Lai JS,Teresi J, Gershon R. Procedures for the analysis of differential item functioning (DIF) for small sample sizes. *Eval Health Prof* 2005;28: 283-294.

33. Zumbo BD. A Handbook on the Theory and Methods of Differential Item Function (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (Ordinal) item Scores. Ottawa. On: Directorate of Human Resources Research and Evaluation, Department of National Defense. 1999.

34. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, No. 17. 1969.

35. Samejima F. Graded response model. In W.J. van der Linden & R.K. Hambleon (Eds.) Handbook of modern item response theory (pp. 85-100). Springer. 1997.

36. Teresi J, Fleishman JA. Differential item functioning and health assessment. *Qual Life Res* 2007;16: 33-42.

37. Demorest ME, Wark DJ, Erdman SA. Development of the screening test for hearing problems. *Am J Audiol* 2011;20:100-110.

38. Duijvestijn JA, Anteunis LJC, Hoek CJ, Van den Brink RHS, Chenault MN, Manni JJ. Help-seeking behavior of hearing-impaired persons aged ≥55 years; effect of complaints, significant others and hearing aid image. *Acta Otolaryngol* 2003;123:846-850.

39. Thissen D. IRTLRDIF v2.0b; Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning. 2001.

40. Crocker LM, Algina J. Introduction to classical & modern test theory (pp145-146). Holt, Rinehart and Winston, Inc. Fort Worth, U.S.A. 1986.

41. Thissen D, Chen WH, Bock RD. MULTILOG 7.0. Chicago: Scientific Software International. 2003.

42. Cai L, Thissen D, du Toit SHC. IRTPRO for Windows [Computer software]. Lincolnwood, IL: Scientific Software International. 2011.

43. Hambleton RK. Good practices for identifying differential item functioning. *Med Care* 2006;44 (11 Suppl. (3): 182-188.

44. Danermark B, Cieza A, Gange JP, Gimigliano F, Granberg S, Hickson L, Swanepoel D. *Intl J Audiol* 2010;49:256-262.

45. Flannery WP, Reise SP, Widaman KF. An item response theory analysis of the general and academic scales of the Self-Description Questionnaire II. J Res Pers 1995;29:168-188.

46. Thomas ML. The value of item response theory in clinical assessment: A review. *Assessment* 2011;18:291-307.

47. Baker FB. *The basics of item response theory*. Portsmouth, NH; Henemann. 1985. Retrieved from http://echo.edres.org.8080/irt/baker/

4

# Chapter 5

Item Response Theory applied to factors affecting the patient journey towards hearing rehabilitation

MN Chenault
MPF Berger
B Kremer
LJC Anteunis

## Abstract

To develop a tool for use in hearing screening and to evaluate the patient journey towards hearing rehabilitation, hearing aid item responses to the Hearing Aid Rehabilitation Questionnaire (HARQ) were evaluated with Item Response Theory. The scales *aid stigma*, *pressure*, and *aid unwanted* address respectively hearing aid stigma, experienced pressure from others, and perceived hearing aid benefit. The sample was comprised of 212 persons aged 55 years or more; 63 were hearing aid users, 64 with and 85 persons without hearing impairment according to guidelines for hearing aid reimbursement in the Netherlands. Bias was investigated relative to hearing aid use and hearing impairment within the Differential Test Functioning (DTF) framework. Items compromising model fit or demonstrating Differential Item Functioning (DIF) were dropped. The *aid stigma* scale was reduced from 6 to 4, the *pressure scale* from 7 to 4, and the *aid unwanted* scale from 5 to 4 items. The resulting scales can be considered bias free ready for screening purposes and application to further understand the help-seeking process of the hearing impaired.

5

## Introduction

Numerous studies have discussed factors which influence whether or not a person with hearing impairment seeks help and whether or not help-seeking results in hearing aid uptake and hearing aid use. Davis et al.[1] reported that the rehabilitation process is initiated on average ten years too late compromising the effectiveness of hearing aid fitting. Hearing screening among adult populations has been proposed to facilitate the process of help-seeking toward rehabilitation and to increase benefit gained from hearing aid fitting.

A description of the hearing impaired patient's journey towards hearing rehabilitation in order to improve hearing screens and interventions is needed.[2] Manchaiah et al.[3] described the patient journey towards rehabilitation as having seven stages: (i) pre-awareness; (ii) awareness; (iii) movement; (iv) diagnostics; (v) rehabilitation; (vi) self-evaluation; and (vii) resolution. The first two stages are relevant to hearing screening as they mark the stages where the hearing impaired person is unaware or in the process of becoming aware of his/her hearing impairment. Since these first two stages are followed by the movement stage where help is sought, it is clear that a hearing screen instrument should primarily target these beginning stages. Meister et al.[4] applied the theoretical framework of the theory of planned behavior to model determinants of help-seeking for hearing problems relative to four of these stages: persons who have noticed they have hearing problems but have not yet sought help (awareness), persons who consulted an ENT specialist/audiologist but had not opted to try a hearing aid (movement); persons who were trying out a hearing aid (rehabilitation) and those who had become hearing aid owners (resolution). Extrinsic motivation, the influence of social pressure or significant others, played a greater role in the first two stages while in the latter two stages it was intrinsic motivation, generated by attitudes and behavioral control, which influenced the intention for rehabilitation.

It would appear then that the focus of the first two stages would be on the hearing experience of the hearing impaired individual. However, there are other factors which influence the transition from (i) being impaired to (v) hearing rehabilitation, which can include interventions besides the most commonly sought one of hearing aid fitting. Seeking help marks the initiation of the rehabilitation process by the hearing impaired person. But whether help is sought is influenced by personality attributes and attitudes of the hearing impaired person. Cox et al.[5] reported that ability to adopt coping strategies for hearing but also cynicism or lack of trust are greater among persons not seeking help than among those who do. Manchaiah's third stage on the patient journey, i.e. the movement stage, is characterized by help-seeking as a result of experienced hearing difficulties. However, factors not directly related to hearing problems play a crucial role in help-seeking behavior. Wu et al.[6] found in their study of persons aged 60 years or more, that willingness to try a hearing aid was not at all

5

related to the degree of hearing impairment suggesting the existence of barriers for the hearing impaired person to enter Manchaiah's movement stage.

Results of a survey performed in the United States revealed that denial, concern about costs and hearing aid stigma were at the base of unwillingness to embark on the road to hearing rehabilitation (National Council on the Aging, 1999).[7] These finding were confirmed by Iacobucci et al.[8] in a study examining the intentions and attitudes of hearing impaired individuals as consumers relative to seeking help and the purchase of a hearing aid. Stigmas attached to hearing problems and hearing aids have been shown to form substantial barriers while the role of significant others can play an encouraging role in the hearing rehabilitation process.[9-11] These factors which are not directly related to experienced hearing impairment may then either encourage or discourage the initiation of hearing aid uptake.

Parette and Scherer[12] discussed the impact of stigma on assistive technologies in general and concluded that stigmas associated with disability and use of assistive technology is integrally related to decisions regarding use. Wallhagen explored the nature of hearing loss and hearing aid stigma in a longitudinal qualitative study.[13] She identified three interrelated experiences, namely alterations in self-perception, ageism and vanity. Alterations in self-perception occur when the individual is confronted with hearing loss and a diminished self-value which is further compounded by the prospect of wearing a hearing aid and the associated stigma of needing one. Closely interrelated to this is how ageism interacts with this perception. The societal ideal is pivoted on vitality and youth and wearing a hearing aid is seen as contradictory to this ideal. Moreover, wearing a hearing aid is avoided for vanity reasons since wearing a hearing aid is viewed as unattractive.

The role of significant others also deserves attention in Manchaiah's patient journey. In a study of a large sample of persons aged 55+ participating in a driving test, Duijvestijn et al. found that the hearing impaired who had sought help had experienced more social pressure to do so than those who had not.[14] The social environment of the individual may act as a catalyst. Family members of hearing impaired persons may become aware of the individual's hearing impairment before he/she does. This occurs in practical situations such as the television being turned on too loud, having to repeat communications to their hearing impaired family member, etc. Often it is the family member who suggests that the hearing impaired person should have their hearing checked. Hickson et al.[15] have identified this as "Third Party Disability", a term included in the World Health Organization (WHO) International Classification of Functioning, Disability and Health (ICF) framework to denote the limitations experienced by family members as a result of the disability of a significant other. Wallhagen in her longitudinal qualitative study also analyzed the views of the partners in her sample of hearing impaired persons and found that negative attitudes towards hearing loss and hearing aids were reinforced by those of the partner.[13]

An additional barrier to hearing aid uptake which has been proposed is a lack in expected benefit and cost. Garstecki and Erler[16] reported that cost was more likely to be reported as a problem in those not opting for a hearing aid than by those who do. However, Amlani et al. reported that even if hearing aids were provided free of charge, 65% of the hearing impaired population would still decline adopting an aid.[17]

Although the focus of a hearing screen should be on the beginning stages of the patient journey, the last two stages of the patient journey also deserve attention. The sixth stage is self-evaluation when the person fitted with a hearing aid either accepts or rejects the hearing aid before entering into the last stage being resolution. Dissatisfaction with hearing aids leads to underuse or non-use by hearing aid owners. Joore et al.[18] reported that 11% of hearing aid users were dissatisfied with their hearing aids. Other studies have reported figures of non-use by hearing aid owners. In a large survey conducted among an older Australian population, 24% of the hearing aid owners reported that they never used their aid.[19] In a follow up on first time hearing aid users, Vuoriahho et al.[20] reported that 36.8% used their aids only occasionally and 5.3% not at all. In a scoping study, which is a form of literature review with the focus on uncovering where there are gaps with possibilities for new research, McCormack and Fortnum[21] explored the reasons why persons who have been fitted for a hearing aid do not use them. They found that the most important reasons for ownership but non-use were related to experiencing lack of benefit, in particular lack of effectiveness in noisy situations, and comfort.

A number of questionnaires has been developed to assess attitudes towards hearing aids. The Expected Consequences of Hearing Aid Ownership (ECHO), a sister questionnaire to the Satisfaction with Amplification with Daily LIFE (SADL) developed by Cox and Alexander,[22] measure hearing aid expectations. These questionnaires have been applied to measure the effect of hearing rehabilitation. Hallam and Brooks[10] developed the Hearing Aid Rehabilitation Questionnaire (HARQ) to improve care for individuals in the rehabilitation process. The HARQ consists of items dispersed over seven scales whereby half of the items pertain to hearing and half to hearing aids. Four scales pertaining to hearing aids were labelled: 'hearing aid stigma', 'pressure to be assessed', 'aid not wanted 'and 'positive expectation of aid'. The original HARQ hearing aid scales have been applied in other studies to assess pre- fitting attitudes: Meister et al.[23] to address expectations and Jerram and Purdy[24] hearing aid stigma.

Chenault et al.[25] analysed the HARQ from the perspective of hearing screening, rather than in the context of hearing rehabilitation, using a sample of persons with and without hearing aids and varying degrees of hearing impairment. Exploratory factor analysis was applied to the hearing aid items from the HARQ scales: 'hearing aid stigma', 'pressure to be assessed' and 'aid not wanted'. The emerging factor structure showed considerable overlap with that reported by Hallam and Brooks[10] but there were still some clear differences. The obtained scales were named: *aid stigma*, *pressure*, and *aid unwanted* targeting respectively attitudes towards the wearing of hearing aids,

5

having experienced social pressure from others to take action regarding hearing impairment, and perceived (lack of) benefit from hearing aids.

The objective of the present paper is to examine these three scales within the methodological framework of Item Response Theory (IRT) to determine their usefulness in a hearing screen setting. Hayes and Lipscomb[26] discussed the advantages of IRT methodology. Not only do IRT models provide a better depiction of actual response patterns, IRT estimates are a more accurate representation of the latent trait being estimated and provide the possibility of gaining sensitivity when comparing groups. IRT is increasingly being applied to Patient Reported Outcomes (PROs) to assess latent traits such as experienced disability and attitudes. IRT is a valuable supplement to Classical Test Theory since it facilitates the further calibration of items and scales, the examination of the relative importance of items within a scale and the investigation of measurement equivalence or bias in responses. To date there have been just a few applications of IRT methodology in audiology studies. Demorest et al.[27] applied IRT methodology to assess items with a pass or fail outcome addressing communication and adjustment to hearing difficulties. The present authors applied this methodology to examine and calibrate two scales, derived from 20 HARQ items pertaining to hearing, to assess experienced hearing.[28,29] The primary goal of the present paper is to evaluate the hearing aid items and their scales in light of the potential contribution they could make to a screening instrument to assess attitudes and factors which either encourage or discourage the initiation of hearing rehabilitation. A secondary goal is to further contribute to the on-going discussion regarding barriers to hearing aid uptake.

## Methods

### Sample

The sample consists of hearing impaired persons with or without a hearing aid and non-hearing impaired persons. The item responses considered in this study were obtained by administering the HARQ to 212 Dutch persons aged 55 and older consisting of 63 hearing aid users and 149 non-users with a mean Pure Tone Average Best Ear (PTABE for 1, 2, 4 kHz) of 38 dB (s.d.16). Hearing aid users were included since a potential screening population will also include persons who have been fitted with a hearing aid and either do or do not use it. Given the criterion for hearing aid reimbursement in the Netherlands at the time of data collection being a PTABE of 35 dB or more for 1, 2 and 4 kHz, the group without a hearing aid could be divided into two groups: non-impaired or normal hearing persons with PTABE less than 35 dB for 1, 2 and 4 kHz in the better ear; $N$=85) and hearing impaired persons with PTABE of 35 dB or more for 1, 2 and 4 kHz in the better ear $N$=64. Hearing aid users were included so that the sample included persons who have gone through the process of hearing aid fitting which includes the realization or acceptance of their hearing impairment.

## HARQ

The hearing scales considered here were obtained by applying exploratory factor analysis to 17 hearing aid items of the HARQ.[25] The same factor structure was obtained through both orthogonal and non-orthogonal rotations resulting in three scales: *aid stigma*, *pressure* and *aid unwanted*. The *aid stigma* scale consists of six items and addresses hearing aid stigma, the *pressure* scale has seven items pertaining to whether the person has been pressured to have his/her hearing assessed, and the *aid unwanted* scale has five items addressing perceived benefit. All items are on a 3-point ordinal response scale: agree, partly agree, and disagree. One item loaded on all three scales and another item on both the *stigma* and *pressure* scales. Reliability analysis resulted in a Cronbach's alpha of 0.62 for the *aid stigma* scale and 0.61 for the *pressure* scale, which are acceptable. For the *aid unwanted* scale a Cronbach's alpha of 0.49 was obtained which is considered unacceptable. However, if this scale had consisted of either 6 or 7 items such as the other two scales considered here, a Cronbach's alpha of 0.55 or 0.58 would have been obtained according to the Spearman Brown Formula, which rounded off would be 0.6 and thus acceptable.

## Scale evaluation

In the Appendix a detailed description of the methodology applied in this study is given. The procedure scale evaluation applied echoes that by Reeve et al.[30] which entails the application of Classical Test Theory for an initial evaluation of the questionnaire scales, determining whether the assumptions required for IRT modelling, namely uni-dimensionality, local independence and monotonicity are met, estimation of the IRT models along with assessment of model and item fit and an inspection of generated response and information curves, followed by Differential Item Functioning (DIF) analysis to ensure that items within each questionnaire scale demonstrate measurement equivalence (are bias free) across groups, before a final re-calibration of the questionnaire scales with the remaining items.

The Classical Test Theory analysis occurred in an earlier paper.[25] Monotonicity was confirmed when item responses were non-monotone decreasing relative to scale scores. Uni-dimensionality was evaluated with Confirmatory Factor Analysis (CFA) according to the methodology of Joreskog,[31] whereby the item coefficients, which are statistically significantly different from zero, and the goodness of fit statistics, Root Mean Square Error of Approximation (RMSEA), the Non-Normed Fit Index (NNFI) and Comparative Fit Index (CFI), were considered. The Graded Response Model (GRM) [32,33] was estimated for each scale, since the items are on a 3-point ordinal scale and the GRM allows the slope parameters to vary per item. The corresponding Item Characteristics Curves (ICC) and Information Curves were generated. Thereafter the obtained IRT model fit measures and parameter estimations were examined. Local independence was evaluated by examining the residual correlations between items

5

generated along with the estimated GRM parameters.[34] Model fit at the item level was assessed by examining the S-$X^2$ per item, an item fit statistic, with $p<0.05$ indicating misfit.[35,36]

The estimated parameters were examined to investigate whether to discard an item from the scale if DIF was detected relative to hearing impairment and hearing aid use. For this purpose three comparisons were considered:

- Comparison I: focal group are normal hearing persons (n=85) with median PTABE (1,2,4 kHz)=25 dB (interquartile range: 18, 28); reference group (n=127) consists of hearing impaired persons with or without a hearing aid.
- Comparison II: focal group are hearing aid users (n=63) with median PTABE(1,2,4 kHz)=51 dB (interquartile range: 42, 62); reference group (149) consists of persons who have either normal hearing or are hearing impaired but not using a hearing aid.
- Comparison III: focal group (n=63) are hearing aid users; reference group (64) consists of hearing impaired persons not using a hearing aid with median PTABE(1,2,4 kHz)=43 dB (interquartile range: 38, 48).

Thus for Comparisons I and II the entire sample is considered. For Comparison III only the hearing impaired persons with or without a hearing aid are considered.

The procedure proposed by Teresi and Fleishmn[37] for the detection of DIF was followed. Two methods, the IRTLR (Item Response Theory Likelihood Ratio) method developed by Thissen et al.,[38] and the OLR (Ordinal Logistic Regression) which is an extension of the LR (Logistic Regression) method developed by Zumbo[39] were used to flag items suspect of exhibiting DIF. These two methods, one being based on IRT calibration and the other on scores obtained through Classical Test Theory sum scores, complement each other.[40] Items free of DIF according to both methods were considered as anchor items to compare the differences in model fit (chi-square statistic) of any item suspect of DIF. This occurred in two stages, first to determine the presence of items exhibiting non-uniform DIF (differences in discrimination parameter estimations across groups) and, secondly, uniform DIF (differences in threshold parameters across groups). Items which were confirmed to exhibit non-uniform DIF were removed from the scale before the procedure was repeated to investigate uniform DIF.

Finally, the (reduced) scales were re-calibrated and the estimated parameters evaluated. Item and total information curves of the (reduced) scales for each of the three scales were examined. Moreover, standardized scale scores derived from the original and reduced scales were compared relative to groups defined by hearing impairment and hearing aid use.

It is important to note that this is a refinement rather than a validation of the factor structure obtained and reported in Chenault et al.,[25] and therefore the results are based on the same sample.

Statistical software employed included LISREL 8.7,[41] IRTPRO Student version and IBM SPSS 21.[36] A type I error of .05 was considered statistically significant. A Bonferroni adjustment for multiple comparisons was applied in the DIF analysis for both the determination of the anchor items and when applying the IRTLR method to suspect DIF items with the subset of anchor items.

## Results

### *Aid stigma* scale

All the factor loadings in the CFA for the six items in the *aid stigma* scale were highly significant supporting the assumption of uni-dimensionality. The mean trait scores per successive response category were non-decreasing. The RMSEA was barely acceptable at 0.0967 with the NNFI at 0.861 and CFI at 0.917. Residual correlations between items ranged from -1 and 1.5 giving evidence of local independence. The estimated slope or discrimination parameters for the six items ranged from 1.1 to 2.8, with the $S$-$\chi^2$ diagnostics indicating that none of the items compromise scale fit. For each item, the content, estimated discrimination parameter $a$ and threshold parameters $b_1$ and $b_2$ are presented in the full scale section on the left in Table 5.1 together with the significance levels of the item fit statistics ($S$-$\chi^2$). In the ICC curves presented in Figure 5.1 it can be seen that for items *as1* and *as2* of the *aid stigma* the peak of the middle curve occurs at higher trait values. For item *as3* this occurs at lower trait values indicating that responders with lower stigma levels are more likely to agree with this statement. The discrimination parameters for all six items are good all being greater than 1. The item information curves are presented in the left top panel of Figure 5.2 showing the relative contribution of each item at each theta trait value. The item with the highest discrimination parameter, *as5*, has correspondingly the highest information peak. It can also be seen that items vary according to the information they provide relative to theta trait values. Item *as3* provides relatively more information at the lower end while item *as2* at the higher end of the trait scale.

In Table 5.2 the chi-square statistics of the DIF analysis are given per scale and comparison. Item *as1* was flagged for non-uniform DIF by the IRTLR method for Comparisons I and II while the OLR method flagged items *as1* for Comparison II and *as3* for Comparisons II and III. Taking items *as2, as4, as5, as6* as anchor items, non-uniform DIF was confirmed for *as1* for Comparisons I and II. Item *as2* was flagged for uniform DIF by the IRTLR method for Comparisons I and II leaving the remaining four items to be considered as anchor items thus confirming uniform DIF for this item for these two comparisons.

5

Table 5.1 Initial and *final scale* estimated IRT parameters (se), fit and scale statistics of *aid stigma, pressure,* and *aid unwanted* scales with reasons for dropping items.

| GRM estimates | Full scale (RMSEA=.097, NNFI = 0.861, CFI=0.917) | | | | Reduced scale (RMSEA=.092, NNFI=0.961, CFI=0.973) | | | |
|---|---|---|---|---|---|---|---|---|
| | a | b1 | b2 | S-X² | a | b1 | b2 | S-X² |
| aid stigma | 6 items | | | | 4 items | | | |
| as1 If I wear a hearing aid people will probably think I am a bit stupid. | 1.1 (0.3) | 1.8 (0.4) | 2.7 (0.6) | 0.62 | (non uniform DIF for Comparisons I, II) | | | |
| as2 It would embarrass me to have to wear a hearing aid. | 1.3 (0.4) | 2.5 (0.6) | 4.0 (1.1) | 0.51 | (uniform DIF for Comparisons I, II) | | | |
| as3 I think people react differently to you when you are wearing a hearing aid. | 1.7 (0.4) | 0.6 (0.1) | 1.2 (0.2) | 0.67 | 1.8(0.4) | 0.6 (0.1) | 1.2 (0.2) | 0.08 |
| as4 I would stand out in a crowd wearing a hearing aid. | 1.1 (0.3) | 1.1 (0.2) | 1.9 (0.4) | 0.48 | 1.1 (0.3) | 1.1 (0.3) | 2.0 (0.4) | 0.59 |
| as5 I think that if you wear a hearing aid people tend to ignore you. | 2.8 (0.9) | 1.1 (0.2) | 1.8 (0.2) | 0.15 | 4.3 (1.7) | 1.0 (0.1) | 1.6 (0.2) | 0.15 |
| as6 It would make me feel old to wear a hearing aid. | 1.5 (0.4) | 1.5 (0.3) | 1.7 (0.3) | 0.71 | 1.2 (0.3) | 1.7 (0.4) | 1.9 (0.4) | 0.05 |
| pressure | 5 items (RMSEA=.078, NNFI =0.94, CFI=0.96) | | | | 4 items (RMSEA=.000, NNFI = 0.98, CFI=1.000) | | | |
| p1 If I wear a hearing aid people will probably think I am a bit stupid. | (local independence assumption violated) | | | | | | | |
| p2 It would embarrass me to have to wear a hearing aid. | 0.6 ((0.4) | 4.5 (2.5) | 7.4 (4.3) | 0.18 | 0.8 (0.4) | 3.3 (1.5) | 5.5 (2.6) | 0.58 |
| p3 I don't consider it important to be assessed for a hearing aid. (reversed) | 1.1 (0.4) | -1.5 (0.4) | -1.2 (0.3) | 0.39 | (uniform DIF Comparison I) | | | |
| p4 My hearing is not so bad that I need a hearing aid. (reversed) | 2.7 (1.4) | 0.4 (0.1) | 0.3 (0.1) | 0.14 | 2.1 (1.0) | 0.5 (0.1) | 0.6 (0.2) | 0.10 |
| p5 I feel I have been pressured into having my hearing assessed. | (unacceptable S-χ² item fit statistic) | | | | | | | |
| p6 I have come here about my hearing in order to please someone else. | 1.5 (0.4) | 1.4 (0.3) | 2.4 (0.5) | 0.73 | 1.6 (0.5) | 1.4 (0.3) | 2.3 (0.5) | 0.62 |
| p7 It is due to pressure from my family or friends that I am having my hearing assessed. | 1.2 (0.4) | 1.9 (0.4) | 2.6 (0.6) | 0.57 | 1.4 (0.5) | 1.7 (0.4) | 2.3 (0.6) | 0.48 |
| aid unwanted | 5 items (RMSEA=.031, NNFI = 0.96, CFI=0.98) | | | | 4 items (RMSEA=.000, NNFI = 0.99, CFI=0.99) | | | |
| au1 It would embarrass me to have to wear a hearing aid. | 0.9 (0.4) | 3.2 (1.2) | 5.3 (2.1) | 0.50 | (uniform DIF Comparison II) | | | |
| au2 From what I know, hearing aids don't help a great deal. | 1.7 (0.5) | 0.2 (0.1) | 1.8 (0.4) | 0.69 | 1.6 (0.5) | 0.2 (0.1) | 1.8 (0.4) | 0.38 |
| au3 I really don't want a hearing aid. | 1.4 (0.3) | 1.9 (0.4) | 2.7 (0.6) | 0.48 | 1.4 (0.5) | 1.9 (0.4) | 2.7 (0.6) | 0.59 |
| au4 I think a hearing aid would help me when meetings trangers. (reversed) | 1.2 (0.3) | -0.1 (0.2) | 0.7(0.2) | 0.92 | 1.2 (0.4) | -0.1 (0.2) | 0.6 (0.2) | 0.83 |
| au5 I am willing to try a hearing aid but I don't think a hearing aid will be of much help to me. | 0.8 (0.3) | 0.5 (0.3) | 1.7 (0.6) | 0.09 | 0.7 (0.3) | 0.5 (0.3) | 1.8 (0.6) | 0.15 |
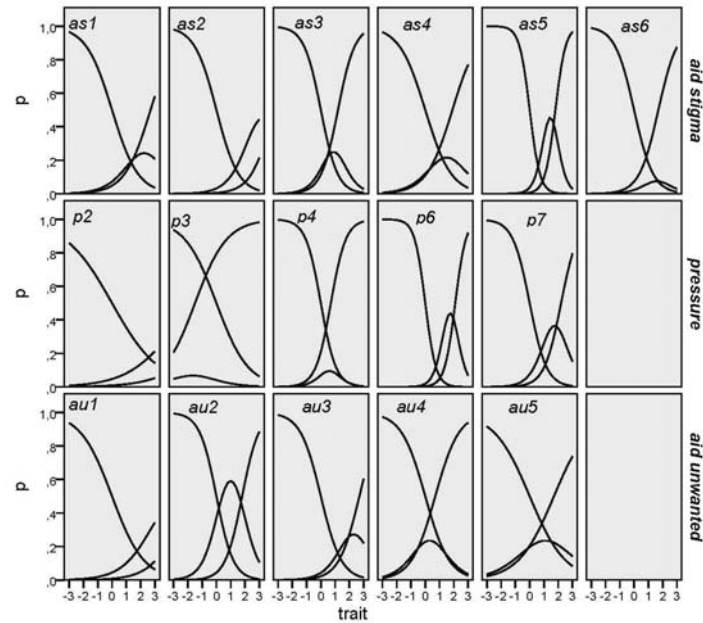
Figure 5.1    Item Characteristic Curves (ICCs) of individual items per scale.
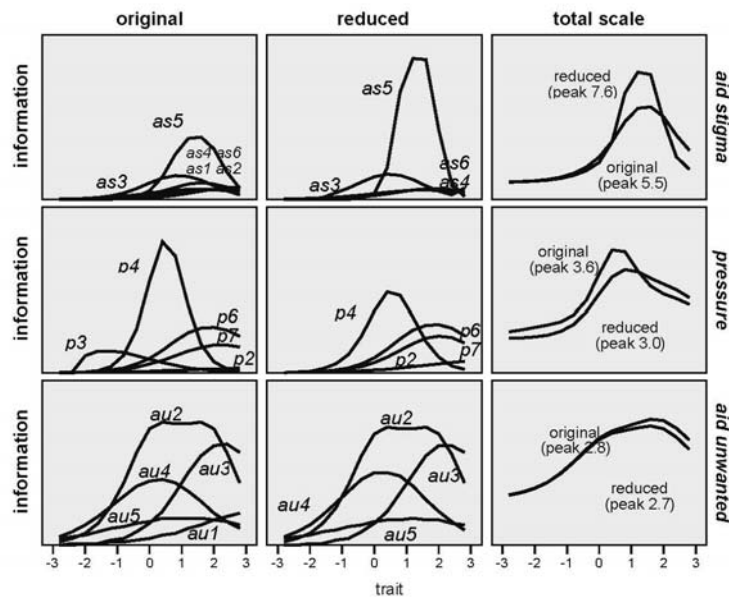


Figure 5.2    Item information curves of each item relative to other items per scale for the original scale on the left, the same for the reduced scale in the center, and a comparison of total information curves for the original and reduced scales.

Table 5.2  $\chi^2$-statistics from DIF analysis for Comparisons I :focal = non-hearing impaired; II: focal is hearing aid group; III: focal=hearing aid, reference=hearing impaired, no aid.

| | | item | Comparison I | | | Comparison II | | | Comparison III | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | IRTLR | OLR | anchor | IRTLR | OLR | anchor | IRTLR | OLR | anchor |
| *aid stigma* | non-uniform | as1 | 8.1* | 2.1 | 12.1 * | 14.9* | 12.5* | 10.5* | 3.6 | 0.5 | 3.0 |
| | | as2 | 4.8 | 0.7 | | 4.7 | 1.2 | | 0.7 | 0.1 | |
| | | as3 | 2.7 | 0.4 | 0 | 0.4 | 8.3* | 3.2 | 2.1 | 12.4* | 1.1 |
| | | as4 | 0.5 | 2.2 | | 1.0 | 3.8 | | 3.3 | 4.7 | |
| | | as5 | 3.6 | 0.9 | | 6.9 | 2.4 | | 1.0 | 0.1 | |
| | | as6 | 0 | 0.5 | | 0.8 | 0.3 | | 1.0 | 0.1 | |
| | uniform | as2 | 7.4* | 5.0* | 7.4*[1] | 7.0 | 4.2 | 7.0*[1] | 2.0 | 0.9 | 2.0 |
| | | as3 | 0.2 | 4.4 | | 1.2 | 4.0 | | 1.7 | 3.4 | |
| | | as4 | 0 | 0 | | 0.3 | 1.2 | | 1.5 | 3.4 | |
| | | as5 | 0 | 1.1 | | 0 | 2.5 | | 1.8 | 0.6 | |
| | | as6 | 0.3 | 0.1 | | 0.6 | 0.1 | | 0.2 | 1.1 | |
| *pressure* | non-uniform | p2 | 0.3 | 0.1 | | 0 | 0 | | 1.2 | 0 | |
| | | p3 | 5.0 | 7.0* | 5.0 | 0.4 | 17.8* | 0.4 | 0.2 | 7.5* | 0.2 |
| | | p4 | 3.0 | 0.1 | | 1.3 | 1.0 | | 0 | 2.4 | |
| | | p6 | 2.7 | 4.5 | | 0 | 0.3 | | 0.1 | 0 | |
| | | p7 | 0 | 2.3 | | 0.8 | 0.1 | | 2.2 | 0.7 | |
| | uniform | p2 | 0.3 | 0.5 | | 0.6 | 0 | | 0.8 | 0 | |
| | | p3 | 9.9* | 11.4* | 9.9* | 2.2 | 6.8 | 2.2 | 2.2 | 5.0 | 2.2 |
| | | p4 | 2.9 | 4.0 | | 2.7 | 2.3 | | 1.6 | 1.3 | |
| | | p6 | 2.7 | 0.1 | | 0.7 | 0.6 | | 0.2 | 0 | |
| | | p7 | 0.2 | 1.1 | | 0.6 | 0.1 | | 0.8 | 0 | |
| *aid unwanted* | non-uniform | au1 | 4.7 | 1.7 | | 1.7 | 0 | | 0.2 | 0 | |
| | | au2 | 0.2 | 2.5 | | 0.9 | 2.6 | | 2.1 | 2.0 | |
| | | au3 | 0.3 | 4.8 | | 0 | 0 | | 1.2 | 0 | |
| | | au4 | 0 | 0.1 | | 0.1 | 2.4 | | 2.7 | 4.4 | |
| | | au5 | 0.6 | 1.3 | | 2.7 | 1.6 | | 0.7 | 4.9 | |
| | uniform | au1 | 4.7 | 0.7 | 5.3 | 14.4* | 11.4* | 13.0* | 8.3 | 5.7 | 6.8 |
| | | au2 | 0 | 5.3 | 0 | 2.8 | 7.0* | 3.7 | 2.7 | 2.9 | 2.6 |
| | | au3 | 0.6 | 0.1 | | 1.0 | 5.9 | | 3.1 | 0.7 | |
| | | au4 | 1.8 | 2.6 | | 3.0 | 2.4 | | 3.7 | 1.1 | |
| | | au5 | 2.7 | 0.1 | | 1.3 | 2.9 | | 3.3 | 0.9 | |

* indicates *p*<.05 after Bonferroni correction; df=1, for non-uniform IRTLR and all OLR comparisons; df=2, for uniform IRTLR comparisons, [1] indicates 1 df for uniform DIF occuring when 0 counts for a response category for one of the groups considered

The assumptions for IRT estimation were evaluated for this reduced four item *aid stigma* scale. Fit statistics were better than for the six item scale. Residual correlations between items were also lower ranging between -0.8 and 0.4. The estimated IRT parameters for the remaining four item scale are given in the reduced scale portion at the right in Table 5.1. Discrimination parameters range from 1.1 to 4.3 with acceptable $S$-$\chi^2$ statistics for all four items. In the top middle panel of Figure 5.2 the relative contribution of each item in the reduced scale is shown. In particular it can be noted that the contribution of item *as5* has increased substantially. In the top right panel of

Figure 5.2 the total information curves generated by the original six and the final reduced four item scales are presented where it can be seen that the reduced scale has a higher maximum peak at 7.6 while the six item scale peaks at 5.5. This is due to the increased contribution of item *as5*. It can also be noted that the reduced scale provides marginally less information at the higher end of the trait values. In Figure 5.3 the standardized scores based on the scales before and after removing items on the basis of item fit and DIF analysis are shown, whereby the diagonal line depicts equality between the two scales. The greatest discrepancy between the two scores is observed in the hearing impaired groups, demonstrating the impact of omitting items *as1* and *as2*.

## *Pressure* scale

The mean trait *pressure* scores per successive response category for each of the seven items were non-decreasing. While the CFA loadings of all seven items were statistically significant, RMSEA was unacceptable at 0.107. Moreover, the residual matrix indicated a high residual correlation of 3.9 between the first two items. These two items are also items in the *aid stigma* scale: "If I wear an aid, people will probably think I'm a bit stupid" and "It would embarrass me to wear a hearing aid". Removing the first of these two items resulted in acceptable values for RMSEA=0.0780, NNFI=0.935 and CFI=0.961. Residual correlations ranged from -0.6 and 1 supporting the assumption of local independence. The $S-\chi^2$ matrix indicated that item *p5* compromised model fit. Removal of this item resulted in acceptable fit statistics and residual correlations ranging from -1.1 to 1.7 and the $S-\chi^2$ statistics indicated acceptable item fit. However the estimated discrimination parameter of item *p2* was low at 0.6 which is visualized in Figure 5.1. The other four items have acceptable discrimination parameters ranging from 1.1 to 2.7. Item *p3* distinguishes itself from the other items by having much lower threshold values with its information being to the left of those of the other four items as can be seen in the middle left panel of Figure 5.2. Item *p4* not only has the highest information peak but appears to provide more information in the middle of the trait scale.

Item *p3* was tagged for non-uniform DIF by the OLR method for all three comparisons. Taking the remaining four items as anchor indicated non-uniform DIF for this item however after adjustment for multiple comparisons this was no longer the case. Item *p3* was also tagged for uniform DIF for Comparison I and the OLR method for all three comparisons with four items again as anchor items. Uniform DIF was confirmed for item *p3* for Comparison I.

Evaluation of the assumptions for IRT estimation for the reduced four item scale indicate even better fit statistics with residual correlations now ranging between -1.4 and 1.4. The IRT model of the four remaining items generated discrimination parameters ranging from 0.8 to 2.1, with acceptable $S-\chi^2$ item fit statistics. In the centre and middle panel of Figure 5.2 the item information curves of the reduced scale show a drop in the peak of item *p4* and a slight increase in item *p2*. In the right middle panel of

this figure the total information curves for the derived four item scale which peaks at 3.0, is shifted to the right of that of the five item scale which peaks at 3.6. In Figure 5.3 it can be seen that the reduced scale has lower standardized scale scores within the hearing aid group and generally higher scores in the non-hearing impaired group.
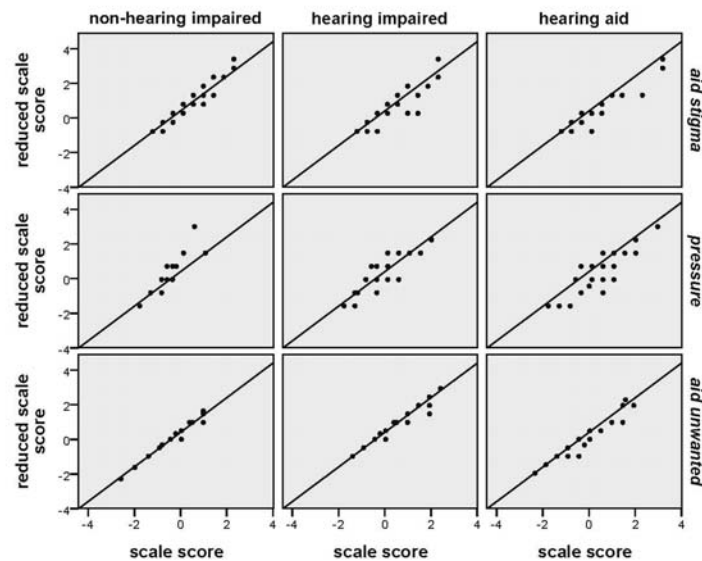


Figure 5.3    Comparison of scale scores for original and final reduced scales for three groups per scale.

### Aid unwanted scale

The CFA loadings of the 5 item *aid unwanted* scale were highly significant while RMSEA was 0.0306, NNFI=0.96 and CFI=0.98. The mean trait scores per successive response category were non-decreasing. Standardized residual correlations ranged from -2.1 to 1.2 in the CFA model while the L-D matrix produced a standardized residual coefficient of 3.3 between items *au2* and *au3* suggesting potential dependency between these two items, but it was decided this was not large enough to remove either item from the scale. Fit at the item level was supported by acceptable $S-\chi^2$ statistics. In the ICC curves presented in Figure 5.1 it can be seen that for items *au1* and *au3* the peak of the middle curve occurs at higher trait values and that *au2* has the highest discrimination parameter. In the bottom left panel of Figure 5.2 it can be seen that item *au2* has correspondingly the highest information peak and moreover that the information provided is much greater than for other items all along the trait scale.

Non-uniform DIF was not detected for any of the five items for each of the comparisons. The IRTLR method tagged *au1* for uniform DIF for Comparison II and the

OLR method items *au1* and *au2* for Comparison II. Subsequent analysis taking item *au3, au4, au5* as anchor items, confirmed uniform DIF for *au1* for Comparison II.

Testing assumptions for the IRT modelling of these four remaining items, indicated excellent fit statistics and residuals ranging from -.007 to 0.007. The IRT model of the four remaining items generated discrimination parameters above 1 except for item *au5* at 0.7. There is little difference in the relative information provided by these remaining four items as can be seen in the bottom middle panel of Figure 5.2. In the bottom right panel it can also be seen that there are negligible differences between the total information curves for the model with 5 items and those obtained for the four item model. In Figure 5.3 it can be seen that the differences between the 5 and 4 item scales are relatively small.

## Discussion

The objective here was to derive scales from the HARQ to assess factors affecting the patient journey towards hearing rehabilitation for use in hearing screening. Three scales, namely *aid stigma*, *pressure* and *aid unwanted,* were evaluated according to fit, item fit and whether the items within each scale demonstrate equivalence relative to experienced hearing aid use and hearing impairment.

The first item of the *aid stigma* scale, "If I wear a hearing aid people will probably think I am stupid", demonstrated non-uniform DIF for Comparisons I (focal group is normal hearing persons) and II (focal group is hearing aid users) and the second item: "It would embarrass me to have to wear a hearing aid" demonstrated uniform DIF for Comparison I (focal group is normal hearing persons). In the original GRM scale calibration these two items had higher threshold values than the other four items in this scale. This means that a positive response to these two items occurs at a higher trait level. In other words, an individual agreeing with these statements experiences relatively more stigma than persons agreeing with the content of the other scale items. When comparing the standardized scores of the six and the reduced four item scales, the largest disparity between the two scores were in the hearing impaired groups with lower scores when the first two items are deleted. Relatively the largest "agree" or "partly agree" responses occur in the hearing aid group, followed by the hearing impaired group with the non-hearing impaired group only responding affirmatively to this item for less than 10% of the cases. The fact that hearing impaired persons respond relatively more affirmatively to this statement possibly illustrates the increased risk of a negative hearing aid image with increasing hearing impairment and it is therefore not a good item to keep in a scale to assess hearing aid stigma, as it appears to be influenced by level of hearing impairment.

Confirmatory factor analysis indicated that the first two items of the seven item *pressure* scale were not locally independent. These two items which are also items in

the *aid stigma* scale are then highly correlated within the *pressure* scale. Removal of the first item improved the fit considerably. IRT calibration of the remaining six items indicated that item *p5*: "I feel I have been pressured into having my hearing assessed" had a highly significant $S$-$\chi^2$ statistic, indicating that removing this item improved the fit of the scale. Item *p5* pertains to experienced pressure but differs from two other items directly pertaining to experienced pressure, namely *p6*: "I have come here about my hearing in order to please someone else" and *p7*: "It is due to pressure from my family or friends that I am having my hearing assessed" in that the words "I feel" in *p5* may add an emotional component.

The DIF analysis was conducted on the scale with items *p2*, *p3*, *p4*, *p6*, and *p7*. Uniform DIF was confirmed for item *p3* when the focal group was the non-hearing impaired group. Lower threshold parameters were obtained for the non-hearing impaired group for this reversed item (*p3*) "I don't consider it important to be assessed for a hearing aid" suggesting that unimpaired persons will interpret this item differently than hearing impaired persons. For non-hearing impaired persons this is most likely an affirmation that they are not experiencing problems while for hearing impaired persons it may indicate a resistance to the pressure they are experiencing from others. This is in line with the study by Duijvestijn et al.[14] where 51% of the non-consulters had reported that others had complained about their hearing. Examination of the total information scales for the five and reduced four item scale shows a shift to the right in information and a slight drop in maximum peak from 3.6 to 3.0.

Toland discusses the usefulness of inspecting information curves to identify items which are redundant.[42] In light of this it is interesting to note how close together the item information curves of *p6* and *p7* are. This of course reflects the similarity in estimated thresholds of these two items, suggesting that these two items could be used interchangeably. Since shorter scales are considered more desirable it could be argued that item *p7* be dropped since item *p6* has the higher discrimination parameter and correspondingly a higher item information curve, being slightly higher all across the trait scale. Since the object of this study was to remove items exhibiting DIF, item *p7* has been retained. If subsequent validation research using another sample indicates that this item is redundant, it should be removed.

The CFA and GRM calibration of the scale *aid unwanted* had very acceptable statistics. The DIF analysis confirmed uniform DIF for the first item, *au1*, when the focal group is the hearing aid user group. This item: "It would embarrass me to have to wear a hearing aid" was also an item in both the *aid stigma* and *pressure* scales. This item had relatively larger estimated threshold values than the other 4 *aid unwanted* items. The DIF analysis indicated that lower threshold values were obtained for the hearing aid group indicating that a person with a hearing aid is more likely to agree with this statement than a person without a hearing aid at the same trait values. Dropping this item, *au1,* from the scale resulted in a scale with acceptable fit statistics and estimated parameters, with only one item having a modest discrimination parameter of 0.7. This

item, *au5*: "I am willing to try a hearing aid but I don't think a hearing aid will be of much help to me" addresses two topics, namely willingness to try a hearing aid and expected benefit. The ambiguity of the focus of this question may in part explain the modest discrimination parameter obtained. It could be questioned whether this item should be maintained in this form. It is ambiguous because it targets two subjects: willingness to try a hearing aid and expected benefit. The first part of this item could be presented as an independent item followed by the second part. Then it would be clear to what extent expected benefit is dependent on willingness. A research area which could be pursued further is to determine latent variables quantifying perceived benefit possibly targeting functional aspects of improving hearing and cost.

It is important to keep in mind the direction of the scales examined here. For the scales *aid stigma* and *aid unwanted*, higher scores reflect more resistance to hearing rehabilitation and thus a greater barrier to hearing aid uptake. Higher scores for the *pressure* scale reflect experiencing more limitations observed by (significant) others which were found by Meister et al.[43] to be the most strongly related to intention for hearing rehabilitation. The *pressure* scale can be viewed as a facilitating factor pushing the hearing impaired individual toward Manchaiah's movement stage. From then on, however, there are intrinsic factors which determine whether rehabilitation is initiated and subsequently successful. The intrinsic factors considered here are hearing aid stigma and perceived benefit. Although they are relevant they are to some extent rooted in personality attributes, which play a role in help-seeking and successful rehabilitation.[5]

There was one scale however in the original HARQ which addressed 'expectation' that presented problems when determining the factor structure in the present sample.[25] These problems could be attributed to these items not being appropriate for our sample of respondents. These items had been formulated for the objectives of the original HARQ to assess expectation at initiation of hearing aid uptake, addressing matters such as how long it would take to become accustomed to an aid. Items which address perceived benefit would be more suitable in a hearing screen such as items in the *aid unwanted* scale like item *au2*: "From what I know, hearing aids don't help a great deal".

It has also been proposed by Wallhagen[13] that routine hearing screening and referral would facilitate valuing hearing loss as a component of overall health. Furthermore it has been suggested that the combination of audiometric measure accompanied by questions regarding experienced hearing problems would filter out those persons who would benefit from hearing rehabilitation. Given that there are other factors not related to experienced hearing which play a role in readiness for hearing aid uptake, a hearing screen should include items or scales which address hearing aid stigma, experienced pressure to initiate hearing rehabilitation and perceived hearing aid benefit. In this paper methods have been applied to obtain scales quantifying these factors.

5

## Conclusion

In this paper we have examined items in scales addressing factors which may either impede or encourage hearing rehabilitation. The items were from the HARQ questionnaire which was originally developed as a tool to improve hearing rehabilitation primarily for counselling purposes. Here questionnaire scales were derived to address factors, other than experienced hearing ability, which play a role in readiness to seek or obtain help for hearing impairment. IRT methodology was applied to reduce scales to only include items which demonstrated measurement equivalence independent of hearing aid use and hearing impairment and also to include items which showed good discriminatory ability along the trait scale for implementation in a hearing screen instrument. A following step is to investigate these refined scales to various levels of hearing impairment and to hearing aid ownership, in an effort to gain understanding regarding the patient journey toward hearing rehabilitation. Moreover, the scales derived and refined here should be validated with another sample.

5

## Appendix

### IRT modelling

Item Response Theory (IRT) is an approach to quantify latent traits. These latent traits are outcomes which cannot be measured directly. In the area of health assessment, questionnaires are increasingly being used to measure patient related traits. The responses to these questionnaires, specifically responses to individual questions or items, are known as Patient Reported Outcomes (PROs). In IRT the items which comprise the questionnaire are examined whereby the characteristics of each item within a scale as well as the scale as a whole are evaluated. The items considered here are on a three point scale of "disagree", "partly agree" and "agree" so that the Graded Response Model (GRM) proposed by Samejima is appropriate.[32,33]

There are two types of parameters which are estimated in IRT analysis. The first is the discrimination parameter of an item with higher estimated discrimination parameters indicating better discriminative ability. The second type of parameter is the threshold parameter. Threshold values indicate the trait value at which transition from one response category into the next occurs, with higher threshold values meaning that relatively higher trait values are required for transition into a higher response category. Since the items considered here have three categories there is one discrimination parameter and two threshold parameters, for the crossover from "disagree" to " partly agree" and from "partly agree" to "agree", to be estimated.

### Testing assumptions for IRT modelling

There are three assumptions required for IRT modelling: monotonicity, uni-dimensionality and local independence as outlined by Edelen and Reeve.[44] Monotonicity, the probability of endorsing one response category higher is non-decreasing with an increase in the latent trait being measured, is investigated by comparing the response counts of successive response categories to scale scores.

Uni-dimensionality can be evaluated with confirmatory factor analysis (CFA) according to the methodology proposed by Joreskog,[31] whereby two aspects are considered: the estimated coefficients of each in the structural equations defining the latent variable and global goodness of fit statistics. For the first, the t-statistics for each of the loading coefficients must be statistically significantly different from zero. The global goodness of fit statistics considered are root mean square error of approximation (RMSEA) as suggested by Shevlin et al.,[45] the Tucker-Lewis or non-normed fit index (NNFI) recommended by Marsh et al.[46] for smaller sample sizes, and the comparative fit index (CFI suggested by Bentler).[47] RMSEA values of less than .05 are ideal and greater than 0.10 poor, while NNFI values above 0.90 are good, greater than 0.95 excellent and CFI values greater than 0.93 desirable.

Local independence (or conditional independence) can be assessed by examining the residual correlations of items within a scale. Chen and Thissen proposed an IRT based test for local independence, which fits the IRT model and produces a LD matrix of residual correlations between items. To assume local dependence these residual correlations must be close to zero.[34]

## Evaluating model fit of the IRT model

Having met the required assumptions, the item parameters can be estimated whereby model fit indices can be obtained for the entire model. The object of modelling is to obtain estimated responses which are as similar as possible to the observed responses. At the same time, a relatively parsimonious model is desired without forfeiting too much information. In IRT modelling this entails comparing the fit of two models, for example, one with a fixed and one with a variable slope or discrimination parameter, since a fixed slope parameter means a reduction in the number of estimated parameters. Chi-square statistics are applied to compare these nested models. Once the fit indices of the model estimated are considered satisfactory, the individual item fit indices can be evaluated. Model fit at the item level can be assessed with the $S\text{-}\chi^2$, an item fit statistic, as suggested by Orlando and Thissen,[35] with $p < 0.05$ indicating that the item does not fit well in the model.

## Model evaluation

Model evaluation, besides obtaining the estimated item discrimination and threshold parameters, also includes inspection of the item characteristic curve (ICC) and information curve for each item in their respective scales. The ICC visualizes each item's discriminative ability and thresholds with higher discrimination or slope values indicating greater discriminatory ability. Discrimination values below 0.8 are considered as poor and values above 1 as acceptable.[48] The first threshold parameter indicates the trait value at which there is a 50% chance of crossing from a "disagree" response to a "partly agree" response and the second from "partly agree" to "agree". The underlying trait being quantified is on the horizontal axis, while the probability of a response is on the vertical axis. As an example, consider Figure 5.4 below where the ICC curves of two items pertaining to hearing aid stigma are shown. Item as3: "I think people react to you differently when you wear a hearing aid" is in the top panel while item as5: "I think that if you wear a hearing aid people tend to ignore you" is in the bottom panel. Each figure shows three curves corresponding to each of the three response categories of "disagree", "partly agree" and "agree". There are two things to note when comparing the ICCs of these two items. First of all, the estimated discrimination parameter of item *as5*, as shown by the relatively sharper peak of the middle curve, is steeper than that for item *as3*. Secondly, the threshold parameters of item *as5* are more to the right than

those for item *as3* indicating that the crossover from a lower to a higher response occurs at relatively higher trait values.
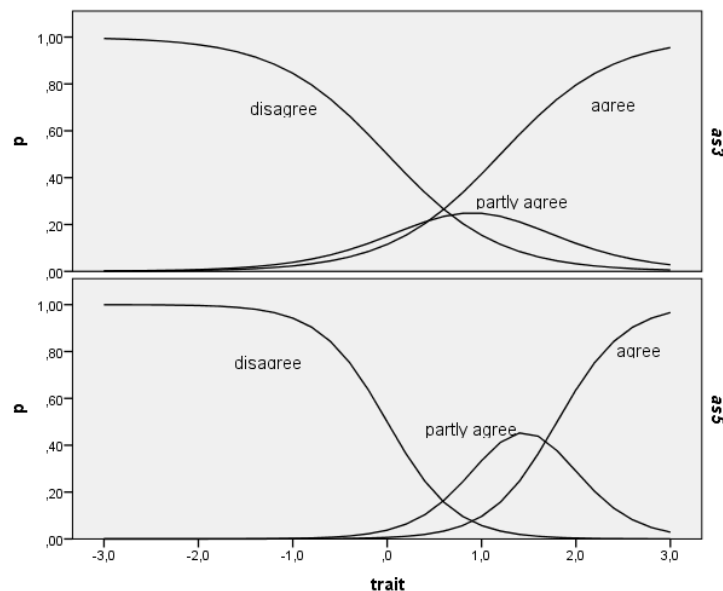


Figure 5.4    Item Characteristic Curve (ICC) of two items, the horizontal axis is a quantification of the underlying trait being measured, the vertical axis is the probability response for each of the response categories ("disagree", "partly agree" or "agree").

The item information curve plots the amount of information provided by an item relative to trait level. When IRT is applied as a diagnostic tool, a steeper curve is more desirable but when persons are to be classified on a wider continuum a peaked information curve may be less useful as the information provided is given in a narrow range of trait values. An information curve with relatively constant values over a wider range may be more useful.[49] Items vary according to the amount of information they provide according to trait levels so that the complementary contribution of items in a scale is relevant. In Figure 5.5 below the item information curves for the same two items are shown. The information curves are generated by the ICC curves. Because the estimated threshold parameters of item as3 are lower than those for as5, the information curve is positioned to the left of that for as5 relative to the axis representing the underlying trait being measured. Also, the information peak of *as3* is lower than that for *as5* which is a result of the less steep discrimination parameter. Relatively more information is provided for trait values above one by *as5* while relatively more information is provided for trait levels below one by *as3*.
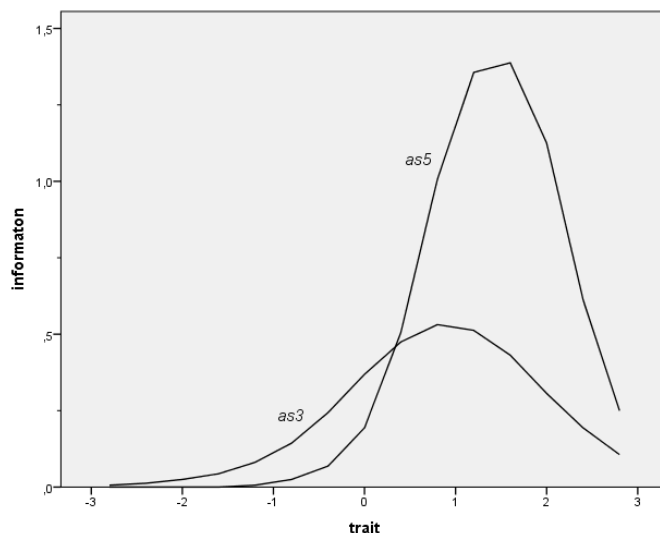
Figure 5.5     Comparison of two information curves.

## DIF methodology

Measurement equivalence is concerned with parameter estimations being invariant across subgroups. When there is measurement equivalence the questionnaire item is interpreted the same way across groups. When the parameters vary across subgroups we speak of differential item functioning so that measurement equivalence cannot be assumed prohibiting a meaningful interpretation of responses or scores across groups. In examining DIF there are three basic elements to consider: (1) item responses, (2) trait level and (3) group membership. It is the effect of group membership on response relative to the underlying trait which determines whether there is DIF or not.

There are various approaches to assess measurement equivalence, but here two methods are applied. One method which relies on IRT calibration is the IRTLR (item response theory likelihood ratio) method whereby the item parameters are estimated separately but simultaneously for the focal group of interest and the reference group while the parameters for other items in the scale are set to be equal across groups. DIF is indicated when there is a statistically significant difference in model fit given by a chi-square statistic. A chi-square statistic exceeding 3.86 is significant for α=.05 when there is one parameter being estimated, and thus 1 degree of freedom, which is the case for the discrimination parameter. The number of threshold parameters is dependent on the number of response categories. If there are three response categories, there are two threshold parameters being estimated, and thus 2 degrees of freedom, so that chi-square statistics exceeding 6 indicate a statistically significant difference in model fit for α=.05. Adjustment for multiple comparisons, if applied, considers the number of items considered in the scale.

The other method is the logistic regression (LR) approach as proposed by Zumbo for items with two response categories.[39] This has been extended for items with more than two ordinal responses to the ordinal logistic regression (OLR) method. This OLR method compares nested ordinal logistic models with as dependent variable the observed item responses. Here the dependent or outcome variable is the response to the item being considered. The first independent variable considered represents a trait scale score which can be estimated by either CTT or IRT methods. In the first step the ordinal response is estimated only from the scale score. The second step includes the group variable in the model for which it is expected that DIF may occur. In the third step the interaction between the trait scale score and group variable is included to predict item response. The coefficients can be interpreted as follows: $\beta_1$ indicates the relation between the observed response to the item relative to the person's scale score (T), while $\beta_2$ is the difference in response between groups (G) conditioned on the scale score and $\beta_3$ is the group by person location interaction. A model fit difference between the second and third models indicates non-uniform DIF, while a model fit difference between the first two models indicates uniform DIF.

An approach for DIF detection and evaluation was proposed by Teresi and Fleishman[37] and consists of six steps: (1) identification of anchor items, anchor items being items which are not suspect of exhibiting DIF, (2) purification of the anchor set, (3) final DIF detection, (4) final parameter estimation and adjustment for multiple comparisons, (5) evaluation of DIF magnitude and (6) evaluation of impact of DIF.

The procedure scale evaluation applied here echoes that by Reeve et al.[30] which entails the application of Classical Test Theory for an initial evaluation of the questionnaire scales, determining whether the assumptions required for IRT modelling, namely uni-dimensionality, local independence and monotonicity are met, estimation of the IRT models along with assessment of model and item fit and an inspection of generated response and information curves, followed by Differential Item Functioning (DIF) analysis to ensure that items within each questionnaire scale demonstrate measurement equivalence (are bias free) across groups, before a final re-calibration of the questionnaire scales with the remaining items.

Helpful background information is available in the form of a text book by Frank Baker (1985) available online: http://echo.edres.org:8080/irt/baker/chapter6.pdf.[50]

# References

1.  Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I. Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models. *Health Technol Assess* 2007;11:1-294.
2.  Smith PA, Davis AC, Pronk M, Stephens D, Kramer SE, Thodi C, Anteunis LJ, Grandori F. Adult hearing Screening: What comes next? (Letter to the Editor) *Int J Audiol* 2011;50:610-612.
3.  Manchaiah VKS, Stephens D, Meredith R. The patient journey of adults with hearing impairment: the patients' views. *Clin Otolaryngol* 2011;36:227-234.
4.  Meister H, Grugel L, Meis M. Intention to use hearing aids: a survey based on the theory of planned behavior. *Patient Prefer Adher* 2014;8:1265-1275.
5.  Cox RM, Alexander GC, Gray GA. Who wants a hearing aid? Personality profiles of hearing aid seekers. *Ear Hear* 2005;26:12-16.
6.  Wu HY, Chin JJ, Tong HMH. Screening for Hearing Impairment in a Cohort of Elderly Patients Attending a Hospital Geriatric Medicine Service. *Singapore Med J* 2004;45:79-84.
7.  Seniors Research Group. The consequences of untreated hearing loss in older patients (Summary). Washington: The National Council on the Aging, May 1999.
8.  Iacobucci D, Calder BJ, Malthouse EC, Cuhachek A. Psychological, marketing, physical, and sociological factors affecting attitudes and behavioural intentions for customers resisting the purchase of an embarrassing product. *Adv Consum Res* 2003;30:236-240.
9.  Hetu R. The stigma attached to hearing impairment. *Scand Audiol Suppl.* 1996;43:12-24.
10. Hallam RS, Brooks DN. Development of the Hearing Attitudes in Rehabilitation Questionnaire (HARQ). *Br J Audiol* 1996;30:199-213.
11. Van den Brink RHS, Wit HP, Kempen GIJM, van Heuvelen MJG. Attitude and help-seeking for hearing impairment. *Br J Audiol* 1996;30:313-324.
12. Parette P, Sherer M. Assistive Technology Use and Stigma. *Educ Train Dev Disabil* 2004;39:217-236.
13. Wallhagen M. The stigma of hearing loss. *Gerontologist* 2009;50:66-74.
14. Duijvestijn JA, Anteunis LJC, Hoek CJ, van den Brink RHS, Chenault MN, Manni JJ. Help-seeking behaviour of hearing impaired persons aged 55 years, effect of complaints, significant others and hearing aid image. *Acta Otolaryngol* 2003;123:836-850.
15. Hickson L, Worall L, Scarinci N. A randomized controlled trial evaluating the active communication education program for older people with hearing impairment. *Ear Hear* 2007;28:212-230.
16. Garstecki DC, J Erler SF. Hearing loss, control and demographic factors influencing hearing aid use among older adults. *J Speech Lang Hear Res* 1998;41:527-537.
17. Amlani AM, Taylor B, Weinberg T. Increasing hearing aid adoption rates through value-based advertising and price unbundling. *Hear Rev* 2011;18:10-17.
18. Joore MA, van der Stel H, Peters HJM, Boas GM, Anteunis LJC. The Cost effectiveness of hearing-aid fitting in the Netherlands. *Arch Otolaryngol Head Neck Surg* 2003;129:297-304.
19. Hartley D, Rochtchina E, Newall P, Golding M, Mitchell P. Use of hearing aid and assistive listening devices in an older Australian population. *J Am Acad Audiol* 2010;21:642-653.
20. Vuorialho A, Sorri M, Nuojua I, Muhli A. Changes in hearing aid use over the past 20 years. *Eur Arch Otorhinolaryngol* 2006;262:355-360.
21. McCormack A, Fortnum H. Why do people fitted with hearing aids not wear them? *Int J Audiol* 2013;52: 360-368.
22. Cox RM, Alexander GC. Expectations about hearing aids and their relation to fitting outcome. *J Am Acad Audiol* 2000;14:403-413.
23. Meister H, Walger M, Brehmer D, von Wedel UC, von Wedel H. The relationship between pre-fitting expectations and willingness to use hearing aids. *Int J Audiol* 2008;47:153-159.
24. Jerram JCK, Purdy SC. Technology, expectations, and adjustment to hearing loss: predictors of hearing aid outcome. *J Am Acad Audiol* 2001;10:445-457.
25. Chenault MN, Anteunis LJC, Berger MPF. Evaluation of the Hearing Aid Rehabilitation Questionnaire in Dutch: examination of its psychometric properties and potential use as a screening instrument. *Audiol Res* 2013;3(e5):32-41.

26. Hays RD, Lipscomb J. Next steps for use of item response theory in the assessment of health outcomes. *Qual Life Res.* 2007;16 Suppl 1:195-199.

27. Demorest ME, Wark DJ, Erdman SA. Development of the screening test for hearing problems. *Am J Audiol* 2011;20:100-110.

28. Chenault MN, Anteunis LJC, Kremer B, Berger MPF. Quantification of esperienced hearing problems with item resonse theory. *Am J Audiol* 2013;22:252-262.

29. Chenault M, Anteunis L, Kremer B, Berger M. An investigation of measurement equivalence in hearing response scales: Refinement of a questionnaire for use in hearing screening. *Am J Audiol* 2015;24: 188-203.

30. Reeve BB, Hays RD, Bjorner JK, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai J, Cella D. Psychometric evaluation and calibration of health-related quality of life item banks. *Med Care* 2007;45:s22-s31.

31. Jöreskog KG. A general approach to confirmatory factor analysis. *Psychometrika* 1969;34:183-202.

32. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement* 1969; No. 17.

33. Samejima F. Graded response model. In W.J. van der Linden & R.K. Hambleton (Eds.) Handbook of modern item response theory (pp85-100). Springer. 1997.

34. Chen WH, Thissen D. Local independence indexes for item pairs using item response theory. *J Educ Behav Stat* 1997;22:265-289.

35. Orlando M, Thissen D. Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *APM* 2013;27:289-298.

36. Cai L, Thissen D, du Toit S. IRTPRO 2.1 for Windows, SSI, Inc. 2012.

37. Teresi JA, Fleishman JA. Differential item functioning and health assessment. *Qual Life Res* 2007;16: 33-42.

38. Thissen D, Steinberg L, Wainer H. Detection of differential item function using the parameters of item response models. In P.W. Holland & H. Wainer (Eds.) Differential item functioning (p 67-113). Hillsdale NJ: Lawrence Erlbaum, Inc. 1993.

39. Zumbo BD. A Handbook on the Theory and Methods of Differential Item Function (DIF): Logistic regression modeling as a Unitary Framework for Binary and Likert-type (Ordinal) item Scores. Ottawa. On: Directorate of Human Resources Research and Evaluation, Department of National Defense. 1999.

40. Lai JS, Teresi J, Gershon R. Procedures for the analysis of differential item functioning (DIF) for small sample sizes. *Eval Health Prof* 2005;28:283-294.

41. Jöreskog KG, Sörbom D. LISREL 8.7 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International, Inc. 2004.

42. Toland MD. Practical guide to conducting an item response theory analysis. *J Early Adolesc* 2014; 34:120-151.

43. Meister H, Walger M, Brehmer D, Von Wedel UC, Von Wedel H. The relationship between pre-fitting expectations and willingness to use hearing aids. *Int J Audiol* 2008;47:153-159.

44. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res* 2007;16:5-18.

45. Shevlin M, Miles JNV, Lewis CA. Reassessing the fit of the confirmatory factor analysis of the multidimensional Students'Life Satisfaction Scale: Comments on "Confirmatory Factor Analysis of the Multidimensional Students'Life Satisfaction Scale." *Pers Individ Diff* 2000;28:181-185.

46. Marsh HW, Balla JR, McDonald RP. Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychol Bull* 1988;103:391-410.

47. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990;107:238-246.

48. Flannery WP, Reise SP, Widaman KF. An item response theory analysis of the general and academic scales of the Self-Description Questionnaire II. *J Res Pers* 1995;29:168-188.

49. Thomas ML. The value of item response theory in clinical assessment: A review. *Assessment* 2011;18: 291-307.

50. Baker FB. *The basics of item response theory*. Portsmouth. NH: Heineman. (online: http://echo.edres.org:8080/irt/baker/chapter6.pdf. 1985.

5

# Chapter 6

Evaluation of the patient journey with the HARQ scales and audiometry

MN Chenault
LJC Anteunis
B Kremer
MPF Berger

## Abstract

**Objectives**

The relation between factors derived from the Hearing Aid Rehabilitation Questionnaire (HARQ), which have been quantified for screening purposes, and hearing impairment classes and hearing aid use were examined to gain insight into the patient journey towards hearing rehabilitation.

**Design**

The sample consisted of 212 persons, 63 of which were hearing aid users, for whom audiometric measures and responses to the HARQ questionnaire were obtained. Discrimination analysis was applied to determine to what extent five HARQ scales: *functionality*, *social hearing*, *pressure*, *aid stigma* and *aid unwanted* quantifying, respectively, disability, handicap, pressure to be assessed, hearing aid stigma and perceived hearing aid benefit, can discriminate hearing aid use and WHO levels of hearing impairment.

**Results**

Disability, as quantified by the HARQ *functionality scale*, appeared to be the best indicator of a person's position on the patient journey. Handicap, pressure from others and hearing aid stigma were good indicators when comparing hearing impaired persons with moderate or mild hearing impairment. Perceived benefit does not appear to be related to hearing impairment but clearly is related to hearing aid use.

**Conclusions**

The items within a scale which distinguish between audiometry outcomes differ according to the level of hearing impairment. This suggests that a tailor made approach to screening be adopted whereby questionnaire items are presented appropriate for the individual's level of impairment.

6

## Introduction

The ageing process is generally accompanied by a gradual deterioration in various aspects of an individual's functionality including hearing ability. Decline in auditory ability with age is generally referred to as presbyacusis, a term which, as pointed out by Kiesling et al.,[1] can be misleading since it infers that age related hearing loss is due to one cause while there are actually numerous causes. Reduction in hearing ability determined through audiometry is known as hearing impairment. In a longitudinal study among older persons, Lee et al.[2] found that average hearing thresholds increased by 1 dB per year for persons aged 60 and more, and that the average rate of change increases with age. While audiograms indicate the level of hearing impairment, making them useful for comparing the prevalence of hearing impairment across populations, they do not quantify the difficulties the individual experiences as a result of his/her hearing impairment. When the individual experiences limitations in hearing functionality we speak of hearing disability. When the individual experiences social barriers or emotional distress as a result of hearing impairment we speak of hearing handicap.

Hearing rehabilitation as an intervention to help the hearing disabled individual improve his/her hearing functionality is customarily seen as synonymous to hearing aid fitting. However, other interventions have been developed to help hearing disabled individuals such as programs to improve communication skills.[3] Clearly the hearing impaired individual must first recognize and subsequently acknowledge the need for rehabilitation, before it can be initiated. Persons, when they are informed of their hearing impairment, often refute the notion that they are hearing impaired.[4] Moreover, after they have become aware of their hearing problems, individuals wait on average ten years before seeking help, while findings suggest that earlier fitting is more beneficial.[5]

Stephens and Kramer[6] suggested that assessment should be based on reported difficulties rather than audiometry. However, reimbursement policy for hearing rehabilitation is generally based on audiometric outcomes, e.g., Pure Tone Averages (PTA) over the 1000, 2000 and 4000 Hz frequencies in either the best or worse ear. For this reason audiometric measurement has often been considered the primary component in hearing screening. Due to the cost of audiometry as part of a screening protocol, a number of studies have investigated alternative approaches to hearing assessment. In a review of screening methods, Yueh et al.[7] concluded that the ideal screening strategy would be a tone-emitting otoscope together with the Hearing Handicap Inventory for the Elderly-Screening Version (HHIE-S) developed by Ventry and Weinstein.[8] Including self-assessment in screening shifts the focus from objective audiometric measures to experienced hearing difficulties.

Hearing screening has been considered as an intervention to trigger awareness among hearing impaired individuals to achieve rehabilitation at an age when the benefit would

6

be greater. Both online and telephone screening has occurred with a follow-up upon failing the screen.[9,10] In programs to target elderly in their home environment, the combination of screening with rehabilitation services has been offered to increase hearing aid uptake.[11,12] While these screening programs increase the number of persons seeking rehabilitation, the percentage of the hearing impaired who take action, resulting in successful hearing aid use, remains very low and non-use increases among hearing aid owners. Understanding the underlying mechanisms which discourage taking steps towards (successful) rehabilitation continues to challenge professionals in the field of audiology.

Several attempts have been made to apply health belief and behavior models to understand help-seeking for hearing rehabilitation. Saunders et al.[13] applied the health belief model to develop the hearing beliefs questionnaire (HBQ) to relate beliefs to hearing related health behavior. The HBQ framework specifies perceived severity and perceived benefits and barriers which together affect help-seeking behavior. Laplante et al.,[14] in a cross-sectional study, applied the trans-theoretical (stages-of-change) model of health behavior change to modify this model to audiology and to determine the stage of change of 224 persons who had failed a Swedish online hearing screening. The derived adapted model consists of four stages of change: pre-contemplation, contemplation, preparation, and action regarding help-seeking for hearing problems.

Since hearing impairment increases with age, the process of becoming hearing impaired and the ensuing increasing difficulties due to hearing impairment has a temporal nature. The onset of the patient journey toward hearing rehabilitation has at its root hearing impairment, which is linked to the ageing process as well as environmental and genetic factors. Somewhere along the track of ageing, most individuals will experience problems and limitations as a result of presbyacusis. Manchaiah et al.[15] identified 7 stages in the patient journey towards hearing rehabilitation. The first stage labelled "pre-awareness" is where the individual may be experiencing hearing difficulties but does not attribute it to his hearing ability, and where possibly mechanisms such as coping and denial occur. In the second stage, "awareness", the individual has become aware of his/her hearing difficulties while in the third stage, "movement", is where consultation is sought. The following stage, "diagnostics" is where audiometric evaluation occurs followed by "rehabilitation" where an intervention such as hearing aid fitting occurs. After this there are 2 more stages being "self-evaluation" where the individual assesses whether the intervention has been successful and the last stage, "resolution" where the person has successfully completed the rehabilitation process.

Although the theoretical framework presented by this patient journey may provide clues regarding the behavior of a person with hearing difficulties, not all the transition points between stages can be identified by a specific action. Becoming a satisfied hearing aid user would place the individual at the end of the patient journey labelled "resolution". First time hearing aid fitting would mark the beginning of the

"rehabilitation stage". Initial consultation regarding hearing problems would mark the transition to the "movement" stage. In contrast, the onset of both the "pre-awareness" and "awareness" stages cannot be stipulated by a specific occurrence. However, there are various mechanisms such as denial and coping and comments from others regarding one's hearing which affect the level of self-awareness of the hearing impaired individual.

A wide range of questionnaires has been developed to assess experienced disability and/or handicap. Most of the questionnaires applied have been implemented as part of the rehabilitation process. Hallam and Brooks[16] developed the Hearing Aid Rehabilitation Questionnaire (HARQ) as a tool to be used in the rehabilitation process to improve the adjustment to hearing aid use among persons who sought help for their hearing problems. The HARQ was thus originally developed to facilitate the transition on the patient journey from the "rehabilitation" to the final "resolution" stage.

Chenault et al.[17] evaluated the HARQ items for screening purposes taking into account that in addition to experienced disability and handicap, there are other factors either promoting or impeding the individual's transition along the patient journey. These factors include experiencing pressure from others to have one's hearing assessed, hearing aid stigma and perceived hearing aid benefit. Hickson et al.[3] identified pressure as stemming from "Third Party Disability", a term included in the World Health Organization (WHO) International Classification of Functioning, Disability and Health (ICF) framework to denote the limitations experienced by family members as a result of the disability of a significant other. Parette and Scherer[18] concluded that decisions regarding rehabilitation are related to stigmas toward disability and assistive technology. Amlani et al.[19] showed that even when hearing aids are offered free of charge uptake remains low. This suggests that lack of perceived benefit plays a greater role in hearing rehabilitation than cost. A screening instrument such as the HARQ could be considered not only as a tool to quantify disability and handicap but also as a tool to detect restraints to seeking and successfully obtaining rehabilitation. Moreover it could be considered as a tool to map persons on the patient journey.

The focus of questionnaire development is to obtain scales, composed of questionnaire items, quantifying latent variables such as handicap or hearing aid stigma. The quantification of these latent variables generally is represented by an aggregation of the item responses belonging to the same scale. While the various items within a scale will target the same latent construct, individual items may differ regarding the amount of information they generate for various levels of the trait being quantified.

In the present paper the object is to determine to what extent the HARQ scales and their individual items, as modified by Chenault et al.,[20-22] can be applied to explore the relation between audiometric outcomes and the patient journey to further contribute to understanding the process the hearing impaired person goes through.

6

## Materials and methods

### Sample

The item responses to the HARQ questionnaire were obtained by administering the HARQ to a sample of 212 Dutch persons aged 55 and older, consisting of 63 hearing aid users and 149 non-users. All items are on a 3-point ordinal response scale: agree, partly agree, and disagree. Exploratory Factor Analysis (EFA) was applied to the original HARQ items, to obtain HARQ scales adjusted for screening purposes. EFA was performed on the 20 hearing items which resulted in 2 scales which were labelled *"functionality"*, addressing experienced hearing ability and *"social hearing"*, handicap as a result of hearing difficulties. EFA was also performed on 17 of the hearing aid items resulting in 3 scales which were labelled *"pressure"*, addressing pressure from others to have one's hearing assessed, *"aid stigma"*, hearing aid stigma, and *"aid unwanted"*, perceived hearing aid benefit.[17] Subsequently Item Response Theory was applied to calibrate the scales and to investigate whether the scales and the items within the scales were bias free relative to hearing aid use and hearing impairment so that they could be used in a hearing screen setting.[20-22] Removing biased items from the two hearing scales resulted in a 9 item *functionality* scale and a 7 item *social hearing* scale with 2 items loading on both scales. Removing biased items from the hearing aid scales resulted in 4 items each for the *pressure*, *aid stigma* and *aid unwanted* scales. The scale score applied here is the mean of the responses of each participant to the items in said scale.

In addition to hearing aid use, the WHO classes of hearing impairment were considered as marking positions along the patient journey. All 212 participants were grouped according to these classes. The number of persons in each class, defined by average thresholds in the better ear over the frequencies 0.5, 1, 2 and 4 kHz, were 54 in No impairment (≤25 dB), 83 in Mild (>25 and ≤40 dB), 63 in Moderate (>40 and ≤60 dB), 12 in Severe (>60 and ≤80 dB), and 0 in Profound (>80 dB). The distribution of the 63 hearing aid users according to the WHO classes were: 1 in No Impairment, 17 in Mild, 35 in Moderate and 10 in Severe.

### Statistical procedure

A check was done to see whether the same factor structure emerged among the 14 hearing and 12 hearing aid items as had emerged among the 20 hearing and 17 hearing aid items in the original analysis.[17] Scale scores were the mean of the responses to the items within said scale, missing values being replaced by the mean of the other items. Mean and standard deviations of these scales were obtained relative to hearing aid use and the WHO classes of hearing impairment.

Discriminant analysis was applied as an inferential tool to determine to what extent each scale correctly classified audiometric outcomes. Due to heterogeneity in group variances, a non-linear discriminant function was estimated. Following this, all 5 scales

were considered in a stepwise procedure with a *p*-value for the *F* statistic below 0.10 for entry and exceeding 0.15 for removal to determine whether a combination of scales could improve the prediction of each of the outcomes considered. The same stepwise procedure was also applied to the items from each scale separately to see which items (within each scale) provided the best classification results. Following this the items from the various scales which had discriminated within their scales were considered together in the same stepwise procedure to arrive at an optimal combination of items for each outcome. For analyses with individual items, only items were included which had less than 10% missing values for each of the two categories of persons being considered. For each analysis, overall percentage of correctly classified cases is reported together with sensitivity (percentage correctly classified for the outcome considered) and specificity (percentage correctly classified in the non-outcome group), using weighted counts because of unequal group sizes. The groups compared were defined either by hearing aid use or the WHO classification of hearing impairment so that four comparisons between participants were made: persons with hearing aids versus persons without a hearing aid, persons with severe versus persons with moderate impairment, persons with moderate versus persons with mild impairment and persons with mild versus persons with no impairment.

The statistical software program IBM SPSS statistics 21 was used for the analysis.

## Results

The factor loadings of the items obtained in the original analysis and the (further) reduced scales are presented in Table 6.1. The retained hearing scale items yielded a structure similar to those obtained for the original scales. However 2 items (*f2=s1* and *f4=s3*) which originally loaded on both scales now loaded predominately on the *social hearing* scale. On the basis of their content it was decided that these 2 items belong in the *social hearing* scale and therefore were dropped from the *functionality* scale, so that for the analysis a 7 item *functionality* scale is used. The same factor structure obtained for the full scales was obtained for the reduced set of hearing aid items (Table 6.2). Table 6.3 presents the scale score means and standard deviations of the 5 scales relative to hearing aid use and WHO classifications. Mean scale scores for the *functionality* and *social hearing* scales are higher for hearing aid users and also for persons with relatively more hearing impairment. The mean scores for the *pressure* and *aid stigma* scales tend to increase with less hearing impairment while the mean aid unwanted scores are relatively higher for the more impaired.

Table 6.1    Item factor loading of the full and reduced *functionality* (*func*) and *social hearing* (*soc*) scales.

| Item | full scales | | reduced | |
|---|---|---|---|---|
| | func | soc | func | soc |
| *f1* It sometimes depresses me when I cannot follow a conversation. | 0.85 | | 0.81 | |
| *f3* I think I've overcome any hearing difficulties I might have through my own efforts. | 0.50 | | 0.78 | |
| *f5* In a conversational group I keep quiet for fear of saying the wrong thing. | 0.66 | | 0.47 | 0.47 |
| *f6* When several people are chatting, it bothers me that I often lose the thread of the conversation. | 0.75 | | 0.60 | |
| *f8* I can hear well enough when I really concentrate. *(reversed)* | 0.75 | | 0.53 | |
| *f9* By and large I am able to hear without difficulty. *(reversed)* | 0.86 | | 0.69 | |
| *f10* My hearing problems are really quite minor. *(reversed)* | 0.70 | | 0.57 | |
| *s1 (=f2)* I dread meeting new people since becoming hearing impaired. | 0.47 | 0.59 | *(0.22)* | 0.75 |
| *s2* My poor hearing sometimes makes me feel really inadequate. | | 0.76 | | 0.90 |
| *s3(=f4)* I find myself avoiding company because conversation is too much effort. | 0.55 | 0.58 | *(0.35)* | 0.72 |
| *s5* As I see it, I am less of a person because of my hearing difficulty. | | 0.65 | | 0.83 |
| *s7* I get the feeling that other people find it a strain to talk to me. | | 0.70 | | 0.73 |
| *s9* My hearing loss makes me feel isolated from other people. | | 0.74 | | 0.86 |
| *s10* I have to admit that deep down I feel restricted by my hearing loss. | | 0.63 | | 0.83 |

Table 6.2    Item factor loading of the full and reduced hearing aid scale items.

| | full scales | | | reduced | | |
|---|---|---|---|---|---|---|
| | aid stigma | pressure | aid unwanted | aid stigma | pressure | aid unwanted |
| *as3* I think people react differently to you when you are wearing a hearing aid. | 0.68 | | | 0.57 | | |
| *as4* I would stand out in a crowd wearing a hearing aid. | 0.79 | | | 0.59 | | |
| *as5* I think that if you wear a hearing aid people tend to ignore you. | 0.68 | | | 0.68 | | |
| *as6* It would make me feel old to wear a hearing aid. | 0.80 | | | 0.82 | | |
| *p2* It would embarrass me to have to wear a hearing aid. | | 0.53 | | | 0.47 | |
| *p4* My hearing is not so bad that I need a hearing aid. (reversed) | | 0.73 | | | 0.85 | |
| *p6* I have come here about my hearing in order to please someone else. | | 0.72 | | | 0.54 | |
| *p7* It is due to pressure from my family or friends that I am having my hearing assessed. | | 0.78 | | | 0.62 | |
| *au2* From what I know, hearing aids don't help a great deal. | | | 0.77 | | | 0.58 |
| *au3* I really don't want a hearing aid. | | | 0.71 | | | 0.56 |
| *au4* I think a hearing aid would help me when meeting strangers. (reversed) | | | 0.72 | | | 0.45 |
| *au5* I am willing to try a hearing aid but I don't think a hearing aid will be of much help to me. | | | 0.58 | | | 0.57 |

Table 6.3    Mean (SD) of scale scores according to hearing aid use and WHO hearing impairment classes. Higher scores indicate more problems.

|  |  | functionality | social hearing | pressure | aid stigma | aid unwanted |
|---|---|---|---|---|---|---|
| Hearing aid | Users (n=63) | 2.2 (0.5) | 1.4 (0.5) | 1.3 (0.4) | 1.3 (0.4) | 1.6 (0.3) |
|  | Non-users (n=149) | 1.5 (0.5) | 1.2 (0.4) | 1.5 (0.2) | 1.4 (0.5) | 1.6 (0.4) |
| WHO | Severe (n=12) | 2.4 (0.5) | 1.8 (0.8) | 1.4 (0.4) | 1.4 (0.5) | 1.7 (0.3) |
| Class | Moderate (n=63) | 2.0 (0.5) | 1.5 (0.6) | 1.4 (0.3) | 1.3 (0.5) | 1.7 (0.3) |
|  | Mild (n=83) | 1.6 (0.5) | 1.2 (0.3) | 1.5 (0.3) | 1.3 (0.4) | 1.6 (0.3) |
|  | None (n=54) | 1.4 (0.4) | 1.2 (0.4) | 1.6 (0.2) | 1.5 (0.6) | 1.6 (0.4) |

In Table 6.4 the correctly classified percentages for hearing aid use and adjacent comparisons of the four WHO classifications according to the five scales are presented. For the classification hearing aid use versus non-hearing aid use, all the scales had reasonably high overall classification percentages. Only *functionality* had a high sensitivity at 87.2%. The scales *aid stigma* and *aid unwanted* had a sensitivity of 0% while *social hearing* and *pressure* had a sensitivity of 14.8% and 42.5% respectively. The stepwise procedure indicated that *functionality*, *pressure* and *aid unwanted* together correctly classified 81.4% of the cases, with a sensitivity of 59.0% and a specificity of 90.6%. This does not appear to be a great improvement above only considering the *functionality* scale for total classification. The difference, however, lies in the sensitivity and specificity values. While sensitivity has dropped from 87.2% to 59.0%, specificity has risen from 54.1% to 90.6% indicating that the scales combined classify the non-hearing aid users very well.

When comparing the severe and moderate hearing impaired groups, overall classification percentages were high with high specificity values and zero or negligible sensitivity values. Only the *functionality* scale emerged as a predictor from the stepwise procedure.

The correct classification percentages when comparing the moderate and mild hearing impaired groups were lower, with specificity generally still higher than sensitivity values. The stepwise procedure resulted in *functionality*, *pressure* and *aid unwanted* together classifying correctly 75.5% of the cases with a sensitivity of 67.2% and a specificity of 82.1%.

Comparison of the mild with the no impairment group resulted in even lower total correct classification percentages. However the sensitivity values were all moderately high being approximately 60% for all the scales. At the same time, *functionality* had the highest specificity value at 40.7%. The stepwise procedure resulted in the *functionality* and *aid stigma* scales together correctly classify 62.8% of the cases.

6

Table 6.4    Discriminant Analysis with scale scores with percentage correctly classified for overall, sensitivity and specificity (% correctly classified per group) for hearing aid users versus non hearing aid users and adjacent WHO groups.

| Outcome | Correctly classified % | Sensitivity/specificity |
|---|---|---|
| Hearing aid (n=63) vs non hearing aid users (n=149) | | |
| *functionality* | 77.6 | 87.2/54.1 |
| *social hearing* | 68.6 | 14.8/90.6 |
| *pressure* | 78.6 | 42.6/93.3 |
| *aid stigma* | 70.8 | 0/100 |
| *aid unwanted* | 69.3 | 0/100 |
| Model of scales together: | 81.4 | 59.0/90.6 |
| *functionality, pressure, aid unwanted* | | |
| Severe (n=12) versus moderate (n=63) | | |
| *functionality* | 83.6 | 0/100 |
| *social hearing* | 84.9 | 16.7/98.4 |
| *pressure* | 83.6 | 0/100 |
| *aid stigma* | 83.6 | 0/100 |
| *aid unwanted* | 83.6 | 0/100 |
| Model of scales together: *functionality* | 83.6 | 0/100 |
| Moderate (n=63) versus mild (n=83) | | |
| *functionality* | 68.1 | 59.0/74.4 |
| *social hearing* | 63.9 | 27.8/90.4 |
| *pressure* | 73.1 | 31.1/88.0 |
| *aid stigma* | 56.6 | 4.9/95.1 |
| *aid unwanted* | 56.3 | 23.0/80.7 |
| Model of scales together: | 75.5 | 67.2/82.1 |
| *functionality, social hearing, pressure* | | |
| Mild (n=83) versus no impairment (n=54) | | |
| *functionality* | 59.1 | 71.1/40.7 |
| *pressure* | 59.9 | 94.0/7.4 |
| *aid stigma* | 60.6 | 91.6/13.0 |
| *aid unwanted* | 61.8 | 85.4/25.9 |
| Model of scales together: *functionality, aid stigma* | 62.8 | 95.2/13.0 |

Table 6.5 presents the results of the discrimination analysis when only items were considered in the stepwise procedure per scale. When comparing the hearing aid users with non-hearing aid users, 4 items emerged from the *functionality* scale, one each from the *social hearing*, *pressure* and *aid stigma* scales while there were 3 items from the *aid unwanted* scale. The *functionality* items which together correctly classified 76.7% of the cases were *f1*: "It sometimes depresses me when I cannot follow a conversation" and *f6*: "When several people are chatting, it bothers me that I often lose the thread of the conversation" and the reversely coded items *f9*: "By and large I am able to hear without difficulty" and *f10*: "My hearing problems are really quite minor". From the *social hearing* scale emerged item *s3*: "I find myself avoiding company because conversation is too much effort" from the *pressure* scale item *p4*: "My hearing is not so bad that I need a hearing aid" and from the *aid stigma* scale item *as4*: "I would stand out in a crowd wearing a hearing aid". The three items which emerged from the *aid unwanted* scale, correctly classifying 75.7% of the sample were *au2*: "From what I

know hearing aids don't help a great deal", *au3*: "I really don't want a hearing aid" and the reversed item *au4*:"I think a hearing aid would help me when meeting strangers". When the above mentioned items were included together in a stepwise procedure, items *p4* and *au4* together correctly classified 91.5% of the cases with high sensitivity and specificity values, meaning that the correct classification of both groups was high.

Table 6.5     Discriminant Analysis with individual scale items with percentage correctly classified for overall, sensitivity and specificity (% correctly classified per group) for hearing aid users versus non hearing aid users and adjacent WHO groups.

| Outcome | Predictors | Correctly classified % | Sensitivity/specificity |
|---|---|---|---|
| Hearing aid (n=63) versus non hearing aid users (n=149) | | | |
| *functionality* | f1 f6 f9 f10 | 76.7 | 50.0/89.8 |
| *social hearing* | s3 | 70.3 | 24.6/89.2 |
| *pressure* | p4 | 92.3 | 90.2/93.2 |
| *aid stigma* | as4 | 71.0 | 0/100 |
| *aid unwanted* | au2 au3 au4 | 75.7 | 63.8/81.1 |
| Model of scale items together | p4 au4 | 91.5 | 89.8/92.3 |
| Severe (n=12) versus moderate (n=63) | | | |
| *functionality* | f10 | 83.1 | 0/100 |
| *social hearing* | s1 | 80.7 | 0/100 |
| *pressure* | - | | |
| *aid stigma* | - | | |
| *aid unwanted* | - | | |
| Model of scale items together | f10 s1 | 80.3 | 33.3/89.8 |
| Moderate (n=63) versus mild (n=83) | | | |
| *functionality* | f5 | 67.4 | 52.5/78.3 |
| *social hearing* | s3 | 70.6 | 44.3/90.2 |
| *pressure* | p4 p7 | 72.9 | 59.0/83.5 |
| *aid stigma* | - | | |
| *aid unwanted* | - | | |
| Model of scale items together | s1 p4 p7 | 75.5 | 67.2/82.1 |
| Mild (n=83) versus no impairment (n=54) | | | |
| *functionality* | f9 | 50.0 | 22.0/92.6 |
| *social hearing* | - | | |
| *pressure* | p4 | 56.3 | 28.4/98.1 |
| *aid stigma* | as4 | 61.8 | 85.4/25.9 |
| *aid unwanted* | - | | |
| Model of scale items together | p4 | 56.3 | 28.4/98.1 |

-: no items resulted from the stepwise procedure

When comparing the severe with the moderate impaired group, the stepwise procedure resulted in only one item each from the *functionality* and the *social hearing* scale. *Functionality* item *f10*: "My hearing problems are really quite minor" correctly classified 83.1% of the cases with zero sensitivity and 100% specificity, indicating that the severe group was not classified at all, while all of the moderate group were correctly classified. *Social hearing* item *s1*:" I dread meeting new people since becoming hearing impaired" correctly classified 80.7% of the cases, with a sensitivity of

zero and a specificity of 100%. These 2 items together provided a total correct classification percentage of 80.3% which is lower than the total correct classification percentages of the items separately. However, specificity increases from zero to 33.3% when the 2 items are combined.

When the moderate and mild impaired groups were compared, one item each from the *functionality* and *social hearing* scales emerged while there were 2 items from the *pressure* and no items from the *aid unwanted* scales. *Functionality* item *f5*: "It sometimes depresses me when I cannot follow a conversation" classified 67.4% correctly. *Social hearing* item *s3*: "I find myself avoiding company because conversation is too much effort" correctly classified 70.6% of the cases. P*ressure* items *p4*:"My hearing is not so bad that I need a hearing aid" and *p7*: "It is due to pressure from my family or friends that I am having my hearing assessed" together provided a correct classification result of 72.9%. When considering these 4 items together in the stepwise procedure, items *s1*, *p4* and *p7* correctly classified 75.5% of the moderate versus mild impaired (sensitivity = 67.2%, specificity = 82.1%).

The *functionality* scale items, *f3* and *f10*, had more than 10% missing values in the group with no impairment and were therefore not included in the stepwise procedure. The stepwise procedure resulted in item *f9*: "By and large I am able to hear without difficulty" with a correct classification rate of 50% to classify into either mild or no hearing impairment. *Pressure* item *p4*: "My hearing is not so bad that I need a hearing aid" and *aid stigma* item *as4*: "I would stand out in a crowd wearing a hearing aid" correctly classify respectively 56.3% and 61.8% of the cases. Considering these 3 items in the stepwise procedure resulted in only item *p4* as discriminating between mild and no hearing impairment.

## Discussion

For each audiometric outcome the discriminating scales and items varied. It is particularly interesting to note that often a select number of items classified as well as the total scale score. Items within a scale vary according to the amount of information they provide reflecting the differences in their estimated IRT item threshold values. Scale scores are an aggregate of the items within that scale so that information provided by scale scores, while maximizing the amount of information provided over the entire trait continuum, will vary in precision along this continuum. Clearly for each of the outcomes considered, when a scale discriminated well it can be attributed to certain items within that scale. While scale scores quantify the latent constructs of disability, handicap, third party disability, attitudes towards hearing aids regarding stigma and perceived benefit, the items within a scale pinpoint the individual's specific situations or opinions and therefore provide clues regarding the individual's position on the patient journey and suggest areas where improvement in rehabilitation services

offered can be made. It is important to emphasize the fact that while increasing hearing impairment is not synonymous with progression along the patient journey, it is nevertheless the underlying contributor to this progression.

One item from the pressure scale, *p4*: "My hearing is not so bad that I need a hearing aid" had extremely high sensitivity and specificity values when classifying hearing aid users, clearly reflecting the fact that hearing aid users are defined by the fact that they have accepted the need for a hearing aid. Also three items from the *aid unwanted* scale (*au2*: "From what I know, hearing aids don't help a great deal", *au3*: "I really don't want a hearing aid" and the reversed item *au4*: "I think a hearing aid would help me when meeting strangers") resulted in a high total correct classification of 75.7%, with a particularly high specificity at 81.1%. The classification results of these *aid unwanted* items are better than the scale as a whole. Apparently, item *au5*: "I am willing to try a hearing aid but I don't think a hearing aid will be of much help" did not contribute to the discriminating ability of the *aid unwanted* scale. The *functionality* scale delivered the best classification results highlighting the fact that it is experienced disability which is related to hearing aid use. The combined scales which discriminated between hearing aid users and non-users were *functionality*, *pressure* and *aid unwanted* reflecting experienced disability, third party disability and perceived benefit respectively. This result was superseded by a total correct classification at 91.5% by the items, *p4* and *au4* from the *pressure* and *aid unwanted* scales respectively.

When comparing severe and moderate hearing impairment, scales and items resulted in high specificity but low sensitivity rates which may in part be attributable to there being only 12 severely hearing impaired persons in the sample. The *social hearing* scale appears to be the best discriminator, being the only scale classifying with a sensitivity value greater than zero, suggesting that it is handicap which distinguishes between severe and moderate hearing impaired. When considering items, the *functionality* item *f10* together with the *social hearing s1* item gave the best results, indicating that disability and handicap combined distinguish between severely and moderately impaired. The *functionality* scale provided the best results for the cutoff point of moderate hearing impairment with even better classification results provided by just one item, namely *p4* which is about the perceived necessity of a hearing aid.

When the moderate group is compared to the mild group, the *functionality*, *social hearing* and *pressure* scales combined gave the best results. However, it is important here to look at the direction of the scales in the analysis. The *functionality* and *aid unwanted* scales were positively and *pressure* negatively related to moderate hearing impairment when compared to mild hearing impairment. This suggests that pressure is being experienced relatively more in the mild group which might be explained by the fact that a substantial portion of the moderately hearing impaired have a hearing aid and thus experienced pressure would be more a matter of the past than the present. At the same time, there is one item in the *pressure* scale, namely *p2:* "It would embarrass me to have to wear a hearing aid", which loaded on the pressure scale but content wise

6

appears to address stigma, contributing further to this result. When considering the items only *p4* emerged as a discriminating item. Persons agreeing with this item believe their hearing is not poor enough to warrant a hearing aid.

When examining the results of mild versus no hearing impairment, the total scales now all had very high sensitivity and low specificity values for the same reason that the specificity was high and the sensitivity low for the cutoff point between severe and moderate hearing impairment. There were 54 persons with no impairment which is almost one fourth of the sample suggesting that it cannot be attributed to the size of this group. Rather it may indicate that there are experienced hearing problems even in the group of persons who are not considered to be hearing impaired which coincides with the findings in the study performed by Hannula et al.[23] The combined scales which provided the best results were the *functionality* and *aid stigma* scales. When the discriminating items were considered, only *pressure* item *p4* emerged. There were two items in the *functionality* scale with more than 10% missing value in the non-hearing impaired group which can be explained by the content of these items not being appropriate for persons in this group. This issue of missing values due to inappropriate items was addressed by Hyde.[24] In a screening population there will be persons with no impairment but a practical solution would be to simply re-phrase such items, or make them contingent on the question as to whether the person experiences any hearing problems at all. Then follow with, for example, item *f3*: I think I've overcome any hearing difficulties I might have through my own efforts" so that the item makes more sense to the respondent.

It appears that disability, quantified by the *functionality* scale, is the most prominent discriminator and, for the various outcomes and ranges of hearing impairment, combines with different scales to achieve better classification results. For greater impairment it appears that handicap, quantified by the *social hearing* scale, plays a role, while for mild impairment, hearing aid stigma, quantified by the *aid stigma* scale, is prominent. The *pressure* scale appears to play a role for moderate hearing outcomes, possibly reflecting third person disability. In particular, item *p4* which addresses perceived hearing aid need even though it is an element of the *pressure* scale, discriminates several outcomes quite well. This item may reflect denial by persons who are resisting pressure, while for persons who have no hearing problems it may simply be an indication that they are not experiencing hearing problems. When considering items targeting disability, moderate versus mild is classified best by item *f1*: "It sometimes depresses me when I cannot follow a conversation" while severe versus moderate by *f6*: "When several people are chatting, it bothers me that I often lose the thread of the conversation". These results suggest the possibility of ordering items relative to degree of hearing impairment which in turn may be linked to the patient journey towards hearing rehabilitation.

Fayers[25] discusses the use of Clinical Adaptive Testing (CAT) to adapt items within a questionnaire to arrive at a more efficient diagnosis in medical diagnosis, rather than

using scale scores derived from lengthy questionnaires. Applying this approach to hearing screening would result in presenting questionnaire items tailored to the level of hearing problems of the participant. Not only would this be less burdensome for the participant, since it results in less items to respond to, but it also focusses on the areas or factors which are relevant for that particular participant. In such a procedure, disability could be assessed first. Then depending on the level of disability, appropriate questions regarding handicap, or stigmas could be presented. Clearly, further study would be required to explore the usefulness of this approach.

There have been some studies which have evaluated the use of questionnaire items in hearing screening by comparing the responses to audiometric outcomes.[7,26] In the present study, not only reported hearing difficulties but also other factors related to hearing problems and hearing aids have been examined relative to hearing impairment. Most notably hearing aid stigma appears to occur at low levels of impairment suggesting that it may certainly play a role in delaying progression along the patient journey but that in the end it is experienced disability which propels the impaired individual further along the patient journey.

The sample considered here includes 146 (of the total of 212) persons being either mildly or moderately hearing impaired representing persons who may or may not display help-seeking behavior. These 2 groups are distinguished from each other by 3 scales together, one pertaining to disability, one having experienced pressure from others to have one's hearing checked and the third perceived benefit. The items which discriminated between these 2 groups pertained to pressure to be assessed and perceived benefit. It appears then that while disability is the most common determinant of the various outcomes, specifically for this outcome of being a hearing aid user, pressure to be assessed and perceived benefit together make a significant contribution.

If the objective is to develop tools to map persons on the patient journey towards hearing rehabilitation, the above results make a valuable contribution. By linking disability, handicap, pressure from others, hearing aid stigma and perceived benefit to hearing aid use and levels of hearing impairment, the relative importance of these factors along the patient journey are demonstrated. This tailor made approach would hopefully help professionals to adapt screening and rehabilitation protocols.

## Conclusions

Disability, as quantified by the HARQ functionality scale, appears to be the best indicator of a person's position on the patient journey. Handicap, pressure from others and hearing aid stigma are good indicators when comparing hearing impaired persons with moderate or mild hearing impairment. Perceived benefit does not appear to be related to hearing impairment but clearly with hearing aid use. The items within a scale

which distinguish between audiometry outcomes differ according to level of hearing impairment. A tailor made approach to screening is advised.

6

# References

1. Kiessling J, Pichora-Fuller MK, Gatehouse S, Stephens D, Arlinger S, Chisolm T, Davis AC, Erber NP, Hickson L, Holmes A, Rosenhall U, von Wedel H. Candidature for and deliver of audiological services: special needs of older people. *Int J Audiol* 2003;42:2S92-2S101.

2. Lee FS, Matthews LJ, Dubno JR, Mills JH. Longitudinal study of pure tone thresholds in older persons. *Ear Hear* 2005;*26:*1-11.

3. Hickson L, Worrall L, Scarcini NA. A randomized controlled trial evaluating the active communication education program for older people with hearing impairment. *Ear Hear* 2007;28:212-230.

4. Humphrey C, Gilhome Herbst K, Faurqi S. Some Characteristics of the hearing-impaired elderly who do not present themselves for rehabilitation. *Brit J Audiol* 1981;15:25-30.

5. Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I. Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models. *Health Technol Assess* 2007;11:1-294.

6. Stephens D, Kramer SE. Living with hearing difficulties: the process of enablement. Chichester Wiley; 2009.

7. Yueh B, Shapiro N, MacLean CH, Shekelle PG. Screening and management of adult hearing loss in primary care. *JAMA* 2003;289:1976-1985.

8. Ventry IM, Weinstein B.V. (1982). The hearing handicap inventory for the elderly: a new tool. *Ear Hear* 1982;3:128-34.

9. Vlaming MSMG, Kollmeier B, Dreschler, WA,Martin R, Wouters J, Grover B, Mohammadh Y, Houtgast T. HearCom: Hearing in the communication society. *Acta Acustica* 2011;97:175-192.

10. Meyer C, Hickson L, Khan A, Hartley D, Dillon H, Seymour J. Investigation of the actions taken by adults who failed a telephone based hearing screen. *Ear Hear* 2011;32:720-731.

11. Linssen AM, Joore MA, Theunissen EJ, Anteunis LJC. The effects and costs of a hearing screening and rehabilitation program in residential care homes for the elder in the Netherlands. *Am J Audiol* 2013;22:186-189.

12. Gussekloo J, de Bont LE, von Faber M, Eekhof JA, de Laat JA, Hulshof JH, van Dongen E, Westendorp RG. Auditory rehabilitation of older people from the general population. The Leiden 85-plus study. *Br J Gen Pract 2003;53:*536-540.

13. Saunders GH, Frederick MT, Silverman S, Papesh M. Application of the health belief model: Development of the hearing beliefs questionnaire (HBQ) and its associations with hearing health behaviors. *Int J Audiol* 2013;52:558-567.

14. Laplante-Lévesque A, Brännström KJ, Ingo E, Andersson G, Lunner T. Stages of change in adults who have failed an online hearing screening. *Ear Hear* 2015;36:92-101.

15. Manchaiah VKS, Stephens D, Meredith R. The patient journey of adults with hearing impairment: the patients' view. *Clin Otolaryngol* 2011;36:227-234.

16. Hallam RS, Brooks DN. Development of the Hearing Attitudes in Rehabilitation Questionnaire (HARQ). *Brit J Audiol* 1996;30:199-213.

17. Chenault MN, Anteunis LJC, Berger MPF. Evaluation of the Hearing Aid Rehabilitation Questionnaire in Dutch: examination of its psychometric properties and potential use as a screening instrument. *Audiol Res* 2013;3:32-41.

18. Parette P, Sherer M. Assistive Technology Use and Stigma. *Educ Train Dev Disab* 2004;39:217-236.

19. Amlani AM, Taylor B, Weinberg T. Increasing hearing aid adoption rates through value-based advertising and price unbundling. *Hear Rev* 2011;18:13:10

20. Chenault MN, Berger MPF, Kremer B, Anteunis LJC. Quantification of experienced hearing problems with item response theory. *Am J Audiol* 2013;22:252-262.

21. Chenault MN, Anteunis LJC, Kremer B, Berger MPFl. An investigation of measurement equivalence in hearing response scales: refinement of a questionnaire for use in hearing screening. *Am J Audiol* 2015;24:188-203.

22. Chenault MN, Berger MPF, Kremer B., Anteunis LJC. Item response theory applied to factors affecting the patient journey towards hearing rehabilitation. *Submitted*. 2015.

6

23. Hannula S, Bloigu R, Majamaa K, Sorri M, Mäki-Torkko E. Self-reported hearing problems among older adults: prevalence and comparison to measured hearing impairment. *J Am Acad Audiol*. 2011;22: 550-559.
24. Hyde ML. Reasonable psychometric standards for self-report outcome measures in audiological rehabilitation. *Ear Hear* 2000;21:24-36.
25. Fayers PM. Applying item response theory and computer adaptive testing: the challenges for health outcomes assessment. *Qual Life Res* 2007;16:287-194.
26. Salonen J, Johansson R, Karjalainen S, Vahlberg T, Isoaho R. Relationship between self-reported hearing and measured hearing impairment in an elderly population in Finland*. Int J Audiol* 2011;50:297-302.
.

6

# Chapter 7

Discussion, conclusions and valorization

# Discussion

In the foregoing chapters the results of analysis of the Hearing Aid Rehabilitation Questionnaire (HARQ) administered to a sample of persons aged 55 years or more are presented with the main objective being to examine latent factors influencing the patient journey towards hearing rehabilitation. In doing so the HARQ has been evaluated as a tool to assess these factors and the role they play in either encouraging or discouraging movement along the patient journey.

The HARQ questionnaire is first analyzed according to Classical Test Theory in Chapter 2, followed by IRT estimation of the model of the hearing scales in Chapter 3, bias detection in the hearing scales in Chapter 4, IRT estimation of and bias detection in the hearing aid scales in Chapter 5, and an evaluation of the classification potential of the five scales and their items regarding hearing aid use and hearing impairment in Chapter 6. The ordering of the chapters is due to the fact that the premises of subsequent chapters are based on findings from preceding chapters.

In Chapter 2 Exploratory Factor Analysis (EFA) is applied to 20 hearing and 16 hearing aid items of the HARQ questionnaire separately and results in two scales from the hearing items and three scales from the hearing aid items. The two hearing scales which quantify disability and handicap are labeled *functionality* and *social hearing*, respectively. The three hearing aid scales are labeled *aid stigma*, *pressure* and *aid unwanted*, quantifying, respectively, hearing aid stigma, pressure to be assessed and perceived benefit.

The validity of the five scales are evaluated relative to hearing aid use, asymmetry in hearing levels, a visual analogue scale for subjective hearing, and various definitions of hearing impairment. These definitions are 35 dB or more in the better ear (Dutch cutoff for hearing aid reimbursement at the time of data collection) and the World Health Organization (WHO) definitions of hearing impairment: no impairment (≤25 dB), mild impairment (>25 and ≤40 dB), moderate impairment (>40 and ≤60 dB), severe impairment (>60 and ≤80 dB).

*Functionality* scale scores increase monotonically with hearing impairment. There is a positive correlation between the *social hearing* scale scores and hearing impairment for the whole sample. However mean *social hearing* scores are lower for the mild hearing impaired group than for the group with no hearing impairment.

A negative relation between *aid stigma* and impairment is only found in the group without hearing impairment at a level of 35 dB. *Aid stigma* scale scores decrease with increasing hearing loss up to 35 dB suggesting that as an individual experiences more hearing problems, negative attitudes towards hearing aids decline. In contrast, *aid stigma* scale scores increase with level of hearing impairment for persons with hearing impairment exceeding 40 dB. This may reflect the stigmatic experiences of hearing aid users while for moderately impaired persons without a hearing aid it may indicate the barrier to hearing rehabilitation presented by hearing aid stigma. The association

7

between *pressure* and subjective hearing found suggests that significant others contribute to how one rates his/her hearing ability. No association is found between *aid unwanted* and any of the audiometric outcomes considered. This can be interpreted as there being no relation between the two or that this scale, with its low reliability, does not adequately quantify hearing aid benefit.

The derived scales form a valuable tool to assess not only hearing difficulties but also determine whether the hearing impaired individual is receptive towards the possibility of hearing rehabilitation. Moreover, the scales can contribute to the relatively unknown early stages of the patient journey towards hearing rehabilitation.

In Chapter 3, Item Response Theory (IRT) is applied to evaluate the two hearing scales *functionality* and *social hearing*. The Graded Response Model (GRM), with estimates for the discrimination parameter when allowed to vary per item, demonstrate a better fit for each scale than the IRT model when the discrimination parameters is estimated as constant over all the scale items. The estimated discrimination and threshold parameters are compared. The generated Item Characteristic Curves (ICCs) and Item Information curves demonstrate the added value of IRT to scale evaluation, whereby the complementary contribution of items within a scale is shown. Total information curves for the three response categories are examined and compared to the total information curves when the response categories "partly agree" and "agree" are merged. This results in a negligible loss in information. The application of IRT methodology can make an important contribution to the assessment of persons in the early stages of the patient journey.

In Chapter 4, methodology from the Differential Test Framework is applied to the two hearing scales *functionality* and *social hearing*. Two items are removed from the *functionality* scale because they exhibit bias in the non-hearing impaired group. Three *social hearing* items are removed, two because they exhibit bias in the non-hearing impaired group and one because it exhibits bias in the hearing aid group. After removal of these biased items the *functionality* scale has nine and the *social hearing* scale seven items. The reduced scales can be considered as being bias free relative to hearing impairment and hearing aid use. This provides evidence that these scales are suitable for further application in a hearing screen setting. An additional advantage is that shorter scales are more convenient for administrative reasons.

There are factors other than experienced hearing difficulties which play a role in readiness for hearing aid uptake. In Chapter 5, the *aid stigma*, *pressure* and *aid unwanted* scales, obtained in Chapter 2 and addressing hearing aid stigma, experienced pressure to initiate hearing rehabilitation and perceived hearing aid benefit, respectively, are further analysed. The IRT models of these three scales are first estimated and evaluated, followed by an investigation of bias using the same protocol outlined in Chapter 4 resulting in in four items in each of these reduced scales.

There is one item, "It would embarrass me to have to wear a hearing aid" which loads on all three scales. Bias is detected in this item for the *aid stigma* scale when the focus

group is the non-hearing impaired group and also in the *aid unwanted* scale when the focus group is the hearing aid group. Bias is not detected for this item in the *pressure* scale. Content wise it would seem that this item belongs in the *aid stigma* scale, but cannot be retained because it exhibits bias in the non-hearing impaired group. This bias in the non-hearing impaired group reflects the higher *aid stigma* scale scores reported in Chapter 2 in this group.

When comparing the standardized *aid stigma* scale scores of the full six and reduced four item scale, the greatest differences occurred in the hearing impaired groups with or without hearing aids. The difference in the two scales scores is due to the removal of the two items demonstrating bias in the non-hearing impaired group. Examination of the responses in the non-hearing impaired group shows that the non-hearing impaired persons are less likely to agree with these two items than with the other four items in the scale, which is the source of the bias. These two items are: "If I wear a hearing aid, people will probably think I am a bit stupid" and "It would embarrass me to have to wear a hearing aid". It may be that non-hearing impaired persons are less likely to agree with these two statements because the notion of wearing a hearing aid is not an approaching reality and therefore are less likely to identify with the emotional content of these items. For the hearing aid group it is the reality while for the hearing impaired group without a hearing aid it may be seen as potentially nearby.

In Chapter 6 EFA is applied to the scale items which were found to be bias free, according to the analyses in Chapters 3 and 4, relative to hearing aid use and level of hearing impairment. The same factor structure can be identified as in Chapter 2, where there were two items loading on both the *functionality* and *social hearing* scales. However, when considering the reduced scales, these two items now load higher on the *social hearing* scale. Since the content of these items clearly address handicap more than disability, it is decided to maintain these items only in the *social hearing* scale. The factor structure determined in Chapter 2 is confirmed in the reduced set of hearing aid items. It is not surprising that these two items loaded on the functionality scale in the initial analysis given the relation between hearing disability and handicap. The removal of biased items from both scales appears to have facilitated a clearer distinction between disability and handicap.

Combinations of scales and items together provide high classification results and remarkably some items together classify the groups better than the scales do. Regarding hearing aid use, the *functionality*, *pressure* and *aid unwanted* together correctly classify almost 59% of the hearing aid users and 91% of the non-users. There were two items, one from the *pressure* and the other from the *aid unwanted* scale which together correctly classify 90% of the hearing aid users and 92% of the non-users. One of these two items, "I think a hearing aid would help me when meeting strangers", appears to reflect the benefit experienced by hearing aid users. An "agree" response to the other item, "My hearing is not so bad that I need a hearing aid", may reflect an attitude of resistance to being pressured to have one's hearing assessed

among persons with hearing impairment. For persons without hearing impairment it may simply reflect the fact that their hearing is indeed good. In any case it is important to realize that an "agree" response to this statement may be interpreted differently by persons with and without hearing impairment. Even though this item was not detected as exhibiting bias, this potential difference in interpretation should be kept in mind when using it in screening.

When comparing adjacent groups according to the WHO classes of hearing impairment, the best classification results are obtained when the moderate and mild hearing impairment groups are compared. The *functionality*, *social hearing* and *pressure* scales together classify 67% of the moderate and 82% of the mild hearing impaired persons. This good classification result is also obtained by three individual items together. One of these items is from the *social hearing* and the other two from the *pressure* scale. The item from the *social hearing* scale is "I dread meeting new people since becoming hearing impaired" while the two items from the *pressure* scale are "My hearing is not so bad that I need a hearing aid" and "It is due to pressure from my family or friends that I am having my hearing assessed". The first of these two *pressure* items may again reflect more than having experienced pressure, in particular, it may indicate the difference in hearing disability between moderate and mild impaired persons.

The relative importance of the factors targeted by the five scales developed along the patient journey is shown. A tailor made approach in assessing an individual's hearing difficulties, experienced pressure from others, attitudes of stigma towards hearing aids and perceived benefit would make a valuable contribution to the adaption of screening and rehabilitation protocols.

In this thesis, factor analysis is performed separately on the HARQ hearing and hearing aid items, just as was originally done by Hallam and Brooks.[1] Items targeting experiencing pressure from others to have one's hearing assessed are included with the hearing aid items. It could be questioned whether these items regarding experiencing pressure from others should not be seen as hearing items rather than hearing aid items. While an individual's subjective experience is unique and personal, it is to a great extent influenced and sometimes formed by his/her environment, which includes persons close to him/her. "Third Party Disability", which has been included in the WHO International Classification of Functioning, Disability and Health (ICF) Framework, pertains to the limitations experienced by family members as a result of the disability of the significant other. If these *pressure* items were included in the factor analysis of the hearing items, the role that third party disability plays in the subjective experience of the hearing impaired person may become more apparent.

Within this context classical test theory is supplemented with methodologies from Item Response Theory (IRT). The added value of IRT is demonstrated by showing that it allows the comparison of items within a scale regarding their discriminative ability and the information they provide relative to the trait levels being quantified by that scale. Moreover, IRT techniques for detecting bias are applied to compare responses of

7

persons with or without a hearing aid and with or without hearing impairment at a level qualifying them for hearing aid reimbursement. This is relevant when designing an instrument for a typical screening population where the range of hearing impairment will be wide and also contains persons who have become unsuccessful hearing aid users. Removal of items which exhibit bias results in a reduction of the number of items per scale which not only has administrative advantages but is less cumbersome for the respondent.

The application of IRT to Patient Reported Outcomes (PROs) is relatively new. There are recommendations for protocols to follow but discussions are ongoing as to exactly what methods and statistical tests are optimal. This is echoed by the increasing availability of computer programs facilitating the application of IRT in a variety of settings, accompanied by lack of conformity about the statistical tests and computing algorithms applied. The choice of software applied and the models estimated is in part determined by the response scale of the questionnaire items being analyzed. In this thesis the Graded Response Model (GRM) is applied because the response scale is ordinal.

Another issue to be addressed when applying IRT is how missing values are dealt with. In Classical Test Theory cases with missing values for some of the items may be deleted. Most often, if the percentage of missing values for a particular item is low, the missing value is replaced either with the mean of the other items in the scale or the item's mean sample response. This is sidestepped in IRT calibration since estimations occur by using the case response to the other items in that scale and the other case responses to that item. This is considered one of the advantages of IRT estimation.

Any measurement tool, and therefore an instrument developed for screening purposes, must be bias free. In other words, a measurement tool must exhibit measurement equivalence across groups. There are various methods for determining the presence of bias. Two methods, one being an IRT approach and the other not, are applied in this thesis. It is good procedure to use more than one approach to assess bias.[2] The two methods applied were recommended by Teresi[3] and the procedure adhered to is that proposed by Teresi and Fleishman.[4] By applying two different methods to determine which items potentially exhibit bias, the possibility of overlooking the presence of item bias is reduced.

Sample size is a matter to be considered when estimating IRT models, with the required number of cases increasing with the number of parameters being estimated. The number of parameters depends on whether the discrimination parameter is constant over all the items within a scale or allowed to vary per item and also on the number of threshold parameters being estimated. If the response set of the items is dichotomous, there is just one threshold parameter, but with more response categories, the number of threshold parameters increases accordingly. When assessing measurement equivalence, the number of parameters is dependent on the bias detection method applied. In any case, larger samples are needed for bias detection than for the

estimation of the IRT model. In Chapters 4 and 5, bias was investigated by comparing item responses of two groups, the reference group and the focal group. Three comparisons were made. The third comparison was between hearing aid users and hearing impaired persons without a hearing aid for which there were only 125 cases. This may have resulted in the non-detection of bias simply due to a limited sample size.

## Conclusions

In this thesis it is shown that disability, as targeted by the *functionality* scale, is closely related to audiometric outcomes and hearing aid use. The distinction between impairment and disability becomes clear. It is not impairment but disability or rather experienced difficulties due to hearing impairment which propels the hearing impaired individual along the patient journey. Given that more handicap as targeted by the *social hearing* scale is observed in the group without hearing impairment than in the group with mild hearing impairment illustrates how individuals differ in how they evaluate their hearing in social situations.

The results presented contribute to clarifying the gaps, as discerned by Duijvestijn et al.,[5] between hearing impaired persons who have hearing aids and those who do not and between hearing impaired persons who have sought help and those who have not.

The individual must go through a phase of experiencing hearing difficulties before he/she is ready to take the next step and enter what Manchaiah terms the "movement" stage. Davis et al.[6] reported that individuals initiate hearing rehabilitation on average ten years later than when benefit would be optimal. Our findings support the notion that having experienced pressure from others appears to act as a catalyst to enter this "movement" stage. It is not clear to what extent this pressure increases awareness or simply functions as a trigger to seek help. Pronk et al.[7] suggested that screening could also contribute to a more timely entry into this "movement" stage to optimise the advantages of hearing rehabilitation. In this context hearing screening would function as an intervention to increase awareness among hearing impaired persons regarding their hearing difficulties. It would be expected that screening as an intervention will increase the number of persons presenting themselves for hearing aid uptake. This timelier and thus earlier initiation of hearing aid uptake would in turn increase the total number of years the individual experiences benefit. This potential gain, however, implies a greater burden for society due to more costs being spread over a longer period of time.

Nevertheless, it must not be forgotten that hearing aid uptake has been shown to be cost effective.[8] Additional gain could be gained by using new tools to improve the rehabilitation process since a serious flaw in hearing aid fitting is that it results in a high number of persons who do not use their hearing aid. A loss then occurs at the societal level but also at the level of the individual person. Not only are costs incurred for an

7

unsuccessful product, but the individual's hearing problems have not been successfully resolved.

In the Netherlands, the delivery of care related to hearing rehabilitation is becoming more and more the terrain of hearing aid dispensers. This commercial sector has recognized the potential market emerging with the ageing population. Moreover, there has been a visible surge in competition between hearing aid dispensers. Not only are there regular announcements of free hearing tests through all the media channels, but also personal letters are sent to senior citizens. In recent years, several eyeglass dispensers have also ventured into the hearing aid market. As a result of this increasing commercialization, hearing aids may be seen more and more as a commodity rather than as a health intervention. With hearing dispensers, unlike health care facilities, being motivated by market mechanisms, it may be concluded that a more efficient allocation of resources towards the services provided is in sight.

It is also conceivable that our ageing population will be accompanied by a shift in attitude regarding hearing aids. While hearing aids have often been viewed as making a person look old, they may become more commonplace so that, in the future, wearing hearing aids will be considered as normal as wearing eyeglasses. If this is the case, the individual will perceive less stigma and attitudes, forming barriers to hearing aid uptake. Possibly the costs associated with hearing aid uptake will be considered less of a hurdle by the potential hearing aid user as well. Moreover, the fact that pension age has been increased will increase the need for hearing aids to attain optimal functioning among relatively older employees.

Probably of greater influence on how hearing aid use is perceived, are the advancements in technology and the prominent position that gadgets have come to play in everyday life. Two studies whereby photos of young persons with and without hearing aids were judged by (graduate) students concluded that hearing aid stigma has declined over the past few decades.[9,10] These findings may in part be attributable to the fact today's younger generation has grown up in a world where computers and electronic gadgets are part of everyday life. The attitudes of today's middle aged and older persons may have at their roots the image of their older generation sometimes wearing conspicuous hearing aids.

Hearing aid fitting is a remedy for an existing problem which here is age-related hearing loss. Decades ago hearing loss due to occupational noise pollution was common. The implementation of protective measures has largely countered this problem. However, a recent study addressed the trend of increasing noise pollution from discotheques and MP3-players which is permanently damaging the hearing of younger persons.[11] It might be expected that this will lead to a reduction in the age at which hearing rehabilitation is initiated. In turn this may reduce the ageism associated with hearing aids. At the same time, it points out the need for prevention programs aimed at increasing the awareness among young people of this potential long term loss in quality of life.

The way in which health care is integrated in a society reflects not only the challenges presented by existing health issues and their priorities, but also is subject to the constraints inherent in a feasible allocation of resources. At the same time, we must remember that choices made at the societal as well as the individual level stem from existing cultural values. Within this context, the International Classification for Functioning stipulated by the World Health Organization provides a framework to distinguish between health states. The application of this framework has had an impact on how health care is prioritized and thus positioned within society. Increasingly emphasis is being placed on the individual's perception of his/her health within the context of his daily functioning and social interactions. The distinction between hearing impairment, hearing disability and hearing handicap sheds light on the impact of impairment on the individual's behavior, in turn facilitating a better understanding of the patient journey. The end of the patient journey was termed by Manchaiah[12] as "resolution" meaning that either the individual's hearing problems have been resolved or the person has resigned him/herself to non-rehabilitation. Ideally, the "resolution" phase would be equitable with hearing aid use. Underuse or non-use of hearing aids has been the topic of a number of studies.[13,14] It has been shown that targeting the non-users among hearing aid owners would be cost-effective. The challenge remains to design hearing rehabilitation protocols which are more successful. Identifying the position of the potential candidate for hearing rehabilitation on the patient journey could reduce waste occurring when persons initiate rehabilitation before they are sufficiently motivated. By applying the scales developed in this thesis, an assessment of motivation or readiness could occur. Once rehabilitation has been initiated, the hearing impaired person should be followed up with supplemental counseling to ensure that the targeted benefit from hearing aid fitting is maximized. The HARQ was originally developed as a tool to improve the rehabilitation process. In this thesis the HARQ has been adapted and refined to assess persons at various stages of the patient journey including stages before initiation of the rehabilitation process.

Recommendations for future research highlight the shortcomings of the research performed here. The results presented are all based on one sample demanding reservation regarding the verity of results obtained. The HARQ should be re-administered in a new sample which includes persons with no hearing impairment and varying levels of hearing impairment , including both hearing aid users and non-hearing aid users. Moreover a group of hearing impaired persons, namely those who have been fitted for a hearing aid but no longer use it, should be included. Administration of our refined HARQ questionnaire to such a group may provide additional insight into why adherence to hearing aid use was not attained. Not including such a group in the present study can be considered a major flaw.

The present study with its cross-sectional design is clearly less ideal than a longitudinal study would be. Since the patient journey is linked to the ageing process, a longitudinal study would provide more insight into how the HARQ items should be ordered on this

7

timeline. Since the sample considered resembles a typical screening population, it can be assumed that a person at each point on the patient journey is represented. It must be recognized, however, that this process is unique for each individual when considering the etiology of hearing impairment, personality attributes which either impede or enhance coping mechanisms, denial and how the individual reacts to his/her social environment. And while age has been shown to influence how the individual rates his/her hearing, this has not been investigated here. Research also shows that education and income level influences the individual's efficacy in successful hearing rehabilitation,[15,16] while a recent study by Benova et al.[17] found no relation between these factors and help-seeking behavior for hearing problems. Future research should investigate these factors and their potential bias.

Another major drawback is that the same sample is used for all the analyses. A validation of the scales in a new sample should occur before the scales derived in this study are applied in a clinical or screening setting.

## Valorization

The research performed for this thesis contributes to understanding the individual as he/she incurs impairment and disability, and the barriers and encouraging factors propelling him/her to gain benefit from hearing rehabilitation. Attention to understanding the position of the individual on the patient journey clarifies what care is appropriate, reducing waste of health services, and increasing the effectiveness of health services given. Moreover, the application of methodologies for applying Patient Reported Outcomes provides an example as to how these methodologies can be applied to health questionnaires targeting other health outcomes. The application of IRT to fine tune questionnaires to assess health status should not only help determine more optimal health care but increase the effectiveness of health care provided as well.

# References

1. Hallam RS, Brooks DN. Development of the Hearing Attitudes in Rehabilitation Questionnaire (HARQ). *Br J Audiol* 1996;30:199-213.
2. Lai JS,Teresi J, Gershon R. 2005. Procedures for the analysis of differential item functioning (DIF) for small sample sizes. *Eval Health Prof* 2005;28:283-294.
3. Teresi J. Different approaches to differential item functioning in health applications. Advantages, disadvantages and some neglected topics. *Med Care* 2006;44:152-170.
4. Teresi JA, Fleishman JA. Differential item functioning and health assessment. *Qual Life Res* 2007;16: 33-42.
5. Duijvestijn JA, Anteunis LJ, Hoek CJ, van den Brink RHS, Chenault MN, Manni JJ. Help-seeking behaviour of hearing impaired persons aged 55 years, effect of complaints, significant others and hearing aid image. *Acta Otolaryngol* 2003;123:836-850.
6. Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I. Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models. *Health Technol Assess* 2007;11:1-294.
7. Pronk M, Kramer SE, Davis AC, Stephens D, Smith PA, Thodi C, Anteunis LJ, Parazzini M, Grandori F. Interventions following hearing screening in adults: A systematic descriptive review. *Int J Audiol* 2011;50:594-609.
8. Joore MA, van der Stel H, Peters HJM, Boas GM, Anteunis LJC. The Cost effectiveness of hearing-aid fitting in the Netherlands. *Arch Otolaryngol Head Neck Surg* 2003;129:297-304.
9. Rauterkus E, Palmer C. Hearing aid use is becoming more accepted. *J Am Acad Audiol* 2013;25:893-903.
10. Clucas C, Karira J, St. Claire L. Respect for a young male with and without a hearing aid: A reversal of the "hearing-aid effect" in medical and non-medical students? *Int J Audiol* 2012;51:739-745.
11. Gommers M, Hoekstra J, Engelfriet P, Wilson C, Picavet S. Gehoorschade en geluidsblootstelling in Nederland. RIVM/Ministerie van Volksgezondheid, Welzijn en Sport. 2013.
12. Manchaiah VKS, Stephens D, Meredith R. The patient journey of adults with hearing impairment: the patients' views. *Clin Otolaryngol* 2011;36:227-234.
13. Hartley D, Rochtchina E, Newall P, Golding M, Mitchell P. Use of hearing aid and assistive listening devices in an older Australian population. *J Am Acad Audiol* 2010;21:642-653.
14. Vuorialho A, Sorri M, Nuojua I, Muhli A. Changes in hearing aid use over the past 20 years. *Eur Arch Otorhinolaryngol* 2006;262:355-360.
15. Erdman SA, Demorest ME. Adjustment to hearing impairment II: Audiological and demographic correlates. *J Speech Lang Hear Res* 1998;41:123-136.
16. Garstecki DC, Erler SF. Hearing loss, control, and demographic factors influencing hearing aid use among older adults. *J Speech Lang Hear Res* 1998;42:527-537.
17. Benova L, Grundy E, Ploubidis GB. Socioeconomic position and health-seeking behavior for hearing loss among older adults in England. *J Gerontol* 2015;70:443-452.

.

7

# Summary

## Summary

In this thesis the Hearing Aid Rehabilitation Questionnaire (HARQ) is employed to understand and evaluate persons with hearing impairment and their potential candidacy for successful hearing rehabilitation. The HARQ was originally developed in the UK to improve care for persons who, having recognized their hearing problems, are initiating hearing aid uptake. It is a well-known fact that a large portion of persons who would benefit from a hearing aid do not present themselves for rehabilitation. Because the HARQ is composed of questionnaire items addressing various aspects of hearing and attitudes towards hearing aids and was considered as a potential instrument for screening purposes, it was administered to a sample including persons who were hearing impaired but dit not have a hearing aid. Also included in the sample were both persons who were not hearing impaired and persons who were successful hearing aid users so that comparisons could be made with persons with hearing impairment but not having a hearing aid. The original HARQ was developed in the UK and has seven scales named: *personal distress*, *hearing loss stigma*, *minimization of hearing impairment*, *hearing aid stigma*, *pressure to be assessed*, *aid not wanted* and *positive expectation of aid*. These scales had been obtained by performing exploratory factor analysis in a sample of persons presenting themselves for hearing rehabilitation. There were two main differences between this original study and our study. First of all, we administered a Dutch translated version of the HARQ. Secondly, our sample differs in that persons with and without hearing impairment and persons with hearing aids were included. For these reasons, we performed exploratory analysis to determine what factor structure was present in our sample responses to the HARQ. A factor structure emerged which had some similarities but was different from the original HARQ. In the screening HARQ which we developed, five scales could be discerned which were labelled *functionality*, *social hearing*, *pressure*, *aid stigma*, and *aid unwanted* targeting, respectively, hearing disability, hearing handicap, pressure by others to have one's hearing assessed, finding heairng aids stigmatic, and (lack) of perceived benefit.

Item Response Theory (IRT) is a methodology, originally developed in educational testing, which is being applied more and more to Patient Reported Outcomes (PROs). IRT was applied to evaluate the five scales of our HARQ screening instrument. We were also interested in determining whether responses to the HARQ were biased relative to hearing aid use and/or hearing impairment. This is important since it is essential that a screening tool demonstrate measurement equivalence in any potential screening population, which could include persons without impairment but also persons who have been fitted for a hearing aid but do not use it. The application of IRT facilitated the comparison of items within their respective scales and to investigate potentially biased items. Items which exhibited bias or compromised the model fit of a scale were removed from their respective scales. The reduced scales were subsequently examined

to determine which scales discriminate between hearing aid users and non-users and groups formed by various levels of hearing impairment. Disability, pressure to have one's hearing assessed and perceived benefit together distinguish hearing aid users from non-users. The scales showed the best discrimination ability when comparing persons who are moderately with persons who are mildly hearing impaired with the moderately impaired reporting relatively more disability and handicap. However the persons with mild hearing impairment, after adjustment for disability and handicap, reported experiencing relatively more pressure to have their hearing assessed. This suggests the role that (significant) others play in becoming aware of one's hearing limitations. The relative importance of the factors on the path towards hearing rehabilitation which are targted by these five scales could be shown.

S

# Samenvatting

## Samenvatting

De Hearing Aid Rehabilitation Questionnaire (HARQ) wordt hier gebruikt om de attitudes van personen die slechthorend zijn te evalueren. De HARQ werd oorspronkelijk ontwikkeld in de UK om het proces van hoortoestelaanpassingen te verbeteren bij personen die erkennen dat zij een gehoorprobleem hebben en als gevolg ervan een gehoorapparaat willen aanschaffen. Het is bekend dat een groot deel van de slechthorende populatie die zouden kunnen profiteren van een hoortoestel zich niet voor aanpassing aanmelden. Omdat de HARQ uit items (vragen) bestaat die over verschillende aspecten van slechthorendheid en hoortoestellen gaan, werd het beschouwd als een geschikt instrument voor screening doeleinden. Het werd afgenomen bij een groep personen boven de leeftijd van 55 jaar waarvan de meeste geen hoortoestel hadden. Van de mensen die geen hoortoestel hadden was er een groep bij die voldeed aan de toen geldende norm voor hoortoestelvergoeding. De oorsponkelijke HARQ zoals ontwikkeld in de UK, bestaat uit 7 schalen namelijk: *personal distress*, *hearing loss stigma*, *minimization of hearing impairment, hearing aid stigma, pressure to be assessed, aid not wanted* en *positive expectation of aid*. Deze schalen werden verkregen door exploratory factor analysis uit te voeren, gebruikmakend van een steekproef van personen die zouden beginnen met hoortoestelaanpassing. Het afnemen van de vragenlijst in deze oorspronkelijke UK studie en onze studie verschilde op twee belangrijke punten. Ten eerste gebruikte wij een Nederlandse versie van de HARQ vragenlijst. Ten tweede bestond onze steekproef uit een heel andere populatie omdat het mensen betrof die of al een hoortoestel hadden of mensen die zich niet hadden aangemeld voor hoortoestelaanpassing. Om deze twee redenen konden wij de vragenlijst niet zomaar in zijn oorspronkelijke vorm gebruiken. Daarom paste wij exploratory factor analysis opnieuw toe om de factor structuur te bepalen. Hoewel er overeenkomsten tussen de oorspronkelijke en door ons verkregen factor structuur waren, bleken er ook wezenlijke verschillen te zijn. Onze HARQ had vijf schalen: *functionality, social hearing, pressure, aid stigma,* en *aid unwanted* die betroffen respectievelijk disability, handicap, druk ervaren van anderen om het gehoor te laten testen, stigmatiserende attitudes t.o.v. hoortoestellen, en verwacht voordeel van een hoortoestel.

Item Response Theory (IRT) is een kwantitatieve methode, oorspronkelijk afkomstig uit de wereld van educatieve testen, die steeds meer wordt toegepast op patiënt gerapporteerde uitkomsten (in het Engels: patient reported outcomes met afkorting PROs). IRT werd in dit proefschrift toegepast om onze vijf verkregen HARQ schalen te evalueren. Wij hebben ook onderzocht of de HARQ bias vertoonde t.o.v. slechthorendheid en/of ervaring met een hoortoestel. Dit is belangrijk omdat een screening instrument meet equivalent moet zijn in elke potentiële screening populatie. Een screening populatie zou uit mensen moeten bestaan zonder gehoorproblemen, mensen die weleens een hoortoestel hebben gekregen (maar het niet langer

gebruiken), alsook mensen die slechthorend zijn en die zich nog voor het eerst voor aanpassing moeten aanmelden. Het toepassen van IRT methodologie faciliteerde de mogelijkheid om items binnen schalen te vergelijken en ook te onderzoeken of een item bias vertoonde, zodat deze items uit hun respectievelijke schalen verwijderd konden worden. Elk van de vijf schalen werd gereduceerd tot een subset van items die vrij van bias was. Deze gereduceerde schalen werden vervolgens gebruikt om te bepalen welke schalen het best tussen mensen met en zonder een hoortoestel, en tussen mensen met verschillende niveaus van slechthorendheid, discrimineerden. De schalen discrimineerden het beste tussen mensen met lichte en mensen met een matige slechthorendheid. De relatieve bijdrage van de verschillende factoren weer-gegeven door deze vijf schalen aan de weg naar rehabilitatie kon worden aangetoond.

S

Publications

## Publications

Chenault MN, Anteunis LJC, Kremer B, Berger M. An investigation of measurement equivalence in hearing response scales: refinement of a questionnaire for use in hearing screening. *American Journal of Audiology* 2015;24:188-203.

De Ruiter AM, Debruyne JA, Chenault MN, Francart T, Brokx JP. Amplitude modulation detection and speech recognition in late-implanted prelingually and postlingually deafened cochlear implant users. *Ear and Hearing* Apr 3 2015 [Epub ahead of print).

Arts RA, George, ELJ, Chenault MN, Stokroos RJ. Optimizing intracochlear electrical stimulation to suppress tinnitus. *Ear and Hearing* 2015;36:125-135.

Muhindo R, Okonya JN, Groves S, Chenault M. Predictors of contraceptive adherence among women seeking family planning services at reproductive health Uganda, Mityana Branch. *International Journal of Population Research* 2015;574329.

Bergshoeff VE, Van der Heijden SJA, Haesevoets A, Litjens SGA, Bot FJ, Voogd AC, Chenault MN, Hopman AHN, Schuuring E, Van der Wa JMl, Manni JJ, Ramaekers FCS, Kremer B, Speel EJM. Chromosome instability predicts progression of premalignant lesions of the larynx. *Pathology* 2014;46:216-224.

Susilo AP, Van Dalen J, Chenault M, Scherpbier A. Informed consent and nurses' roles: a survey of Indonesian practitioners. *Nursing Ethics* 2014;21:684-694.

Chenault MN, Berger MPF, Kremer B., Anteunis LJC. Quantification of experienced hearing problems with Item Response Theory. *American Journal of Audiology* 2013;22: 252-262.

Chenault MN, Anteunis LJC, Berger MPF. Evaluation of the Hearing Aid Rehabilitation Questionnaire in Dutch: examination of its psychometric properties and potential use as a screening instrument. *Audiology Research* 2013;3,e5 32-41.

Smit A, Seehase M, Stokroos R, Jellema R, Felipe L, Chenault M, Anteunis L, Kremer B, Kramer B. Functional impairment of the auditory pathway after perinatal asphyxia and the short term effect of antenatal propofol anesthesia in lambs. *Pediatrics Research* 2013;74,3438.

Hof, JR, Stokroos, RJ, Wix E, Chenault MN, Gelders E, Brokx J. Auditory maturation in premature infants, a potential pitfall for early cochlear implantation. *Laryngoscope* 2013;123:2013-2018.

P

Waterval JJ, Van Dongen TMA, Stokroos RJ, De Bondt RBJ, Chenault MN, Manni JJ. Imaging features and progression of hyperostosis cranialis interna. *American Journal of Neuroradiology* 2012;33:453-461.

Lok W, Anteunis LJC, Chenault MN, Meesters C, Haggard MP. Screening for hearing loss versus parental concern because of ear and/or hearing problems and subsequent referral and treatment for otitis media in the Netherlands. *Scandinavian Journal of Primary Health Care* 2012;30: 163-168.

Lok W, Anteunis LJC, Meesters C, Chenault MN, Haggard MP. Risk factors for failing the hearing screen due to otitis media in Dutch infants. *European Archives of Otorhinolaryngol* 2012;269: 2485-2496.

Lok W, Chenault MN, Anteunis LJC, Meesters CMG, Haggard MP. Selecting infants with OM that need referral and further assessment. Creating a case-finding instrument. *International Journal of Pediatric Oto-Rhino-Laryngology* 2011;75;1301-1307.

Ubbink SW, Van Dijk P, de Kleine E, Brienesse P,Chenault MN, Tan F, Anteunis LJC. Frequency Shifts with age in click-evoked otoaccoustic emissions of preterm infants. *Journal of the Accoustical Society of America* 2011;129:3788-3796.

Van de Berg R, Stokroos RJ, Hof JR, Chenault MN. Bone-Anchored Hearing Aid: A Comparison of Surgical Techniques. *Otology & Neurotology* 2010;31:129-135.

Duijvestijn JA, Grutters JPC, Chenault MN, Joore MA, Manni JJ, Anteunis LJC. Shared care for hearing complaints: guideline effects on patient flow. *Journal of Evaluation in Clinical Practice* 2011;17:209-214.

DuMoulin MFMT, Chenault MN, Tan ES, Hamer JPH, Halfens RJG. Effectiveness of quality systems to improve continence care in the homecare setting. *Journal of Clinical Nursing* 2010; 19:23-24.

DuMoulin MFMT, Chenault MN, Tan ES, Hamer JPH, Halfens RJG. Quality systems to improve care in older patients with urinary incontinence receiving home care: do they work? *Quality & Safety in Health Care* 2010; 19:1-7.

Zumach A, Chenault MN, Anteunis LJC, Gerrits E. Speech perception after early-life otitis media with fluctuating hearing loss. *Audiology & Neurotology* 2010;16:304-314.

P

Grutters J, Anteunis LJC,Chenault MN, Joore MA, Willingness to pay for a hearing aid: comparing the payment scale and open-ended question. *Journal of Evaluation in Clinical Practice* 2009;15:91-96.

Lok W, Chenault MN, Anteunis LJC. Implementation of neonatal screening for hearing impairment: influence on pediatric otitis media surgery in the Netherlands. *International Journal of Pediatric Otorhinolaryngology* 2009;73:1090-1094.

Smeets E, Chenault M, Curfs L, Schrander-Stumpel C, Frijns JP. Rett syndrome and long-term disorder profile. *American Journal of Medical Genetics* 2009;Part 1. 149 A: 199-205.

Zumach A, Gerrits E, Chenault MN, Anteunis LJC.  Otitis media and speech-in-noise recognition in school-aged children. *Audiology & Neurotology* 2009;14:121-129.

Zumach A, Gerrits E., Chenault M, Anteunis L. Long-term effects of early otitis media on language development. *Journal of Speech, Language and Hearing Research* 2010;3: 34-43.

Timmerman AA, Anteunis LJC, Chenault MN, Meesters CMG, Haggard MP. Psychometric evaluation of the OM8-30 questionnaire in Dutch children with otitis media. *European Archives of Oto-Rhino-Laryngology and Head & Neck Surgery* 2008;265:1047-1056.

Dammeijer PFM, Van Dijk P, Chenault MN, Manni JJ. Stapedius muscle fibre characterization in the noise exposed and auditory deprived rat. *Hearing Research* 2007;233:54-68.

Timmerman AA, Meesters CMG, Anteunis LJC, Chenault MN. The level of psychosocial adaptation in young school children with otitis media*. International Journal of Pediatric Otorhinolaryngology* 2007;71:1843-1848.

Hof JR, Anteunis LJC, Chenault MN, Van Dijk P. Otoacoustic emissions at compensated middle ear pressure in children. *International Journal of Audiology* 2005;44:317-320.

Hof JR, Van Dijk P, P Chenault MN, Anteunis LJC. A two-step scenario for hearing assessment with otoacoustic emissions at compensated middle ear pressure (in children 1-7 years old). *International Journal of Pediatrics and Otorhinolaryngology* 2005;69:649-55.

P

Duijvestijn JA, Anteunis LJC, Hoek CJ, Van den Brink RHS, Chenault MN, Manni JJ. Help Seeking of the hearing impaired aged 55 years and older; effect of complaints, significant others and hearing aid image. *Acta Otolaryngology* 2003;123:846-850.

Brunnenberg DEM, Chenault MN, Anteunis LJC. Changing characteristics of first time hearing aid applicants in the Netherlands in the past decade. *International Journal of Audiology* 2003;43: 211-217.

Joore MA, Brunenberg DEM, Chenault MN, Anteunis LJC. Societal effects of hearing aid fitting among the moderately hearing impaired. *International Journal of Audiology* 2003;42:152-160.

Dammeijer PFM, Schlundt Bodien QCM, Chenault MN,. Manni JJ, Anteunis LJC. Effects of early auditory deprivation and stimulation on auditory brainstem responses in the rat. *Acta Otolaryngology* 2002;122:703-708.

Engel JAM, Anteunis LJC, Chenault MN, Marres EHMA. Otoscopic findings in relation to tympanometry during infancy. *European Archives of Oto-Rhino-Laryngology* 2000;257:366-371.

Wijnen JAG, Tan ES, Van Baak MA, Bovens AMP, Vrencken JGP, Verstappen FJT, Schim van der Loeff-Chenault MN. Exercise Systolic blood pressure in middle-aged women with normal and elevated blood pressure. *International Journal of Sport Medicine* 1995;16:61-65.

P

Acknowledgements

A

And lastly I am thankful for the many good fortunes I have received, most importantly the joy and challenges provided by my children and their partners and offspring.

A

# Curriculum Vitae

## Curriculum Vitae

Michelene (Mickey) Chenault was born in Walla Walla in the state of Washington, U.S.A., on August 16, 1948. She grew up in Spokane, Washington, where she graduated from John Rogers High School. After obtaining a Bachelor's Degree with a major in economics from Whitman College she worked for several years in international banking, first in Oregon, U.S.A., and thereafter in Rotterdam. She was a stay-at-home mother for eight years during which time she secured a degree to teach high school mathematics and certification for teaching information science. Following a period wherein she taught high school mathematics, she taught computer programmers for the Rotterdam Chamber of Commerce and worked as a programmer and research assistant at the Central Plan Bureau. In September 1988 she started working part-time as a computer software trainer and research assistant at the Department of Methodology and Statistics of the then Rijksuniversiteit Limburg. In this position, her teaching responsibilities were extended. She contributed to various masters programs including the Masters for Health Professionals in Education, Masters of Public Health and the Advanced Masters in Research Psychology. From 2010 until 2014 she was coordinator of the module Advanced Statistics for the Masters in Global Health program. Moreover, she has been involved in two projects for the MUNDO. The first project was in 2009 where she delivered a week long workshop in Hanoi on data quality control to representatives from the various medical faculties in Vietnam. The second project, which is still ongoing, is providing support to improving the educational program to train health statisticians for the Health Information Management program of the Higher Institute of Health Science in Sanaa, Yemen. In 2010 she spent two weeks at the Afhad University for Women in Obdurman, Sudan, teaching data analysis to masters' students and faculty.

In 1997 her half-time job at the Methodology and Statistics department was complemented with a position as statistical analyst at the Ear Nose and Throat Department of the then Academic Hospital Maastricht to provide support specifically for audiological research projects. Involvement in various projects triggered the motivation for pursuing the research presented in this thesis. Some projects have targeted hearing related problems among small children whereby the initial project was the MOMES study with focus on otitis media with effusion in infants. The EHIMA project followed with focus on age-related hearing. Working on this project formed the beginning of the present study. With a need for more advanced methodological expertise, she enrolled in the Quantitative Analysis in the Social Sciences advanced master's program at the Catholic University Brussels and completed the program (cum laude) in 2009.

Most importantly, Mickey is the proud mother of three adult children: Jenny, Hans and Berra, and also the happy grandmother of three delightful girls: Kiki, Puck and Keet. A fourth grandchild is expected to be born in March.

CV