Item response theory applied to factors affecting the patient journey towards hearing rehabilitation

Michelene Chenault,[1-3] Martijn Berger,[4] Bernd Kremer,[1,5] Lucien Anteunis[1,5]

[1]Department of Otorhinolaryngology and Head and Neck Surgery, Maastricht University Medical Centre, Maastricht; [2]Department of Health, Ethics and Society, Medicine and Life Sciences, Maastricht University, Maastricht; [3]CAPHRI (School for Public Health and Primary Care), Medicine and Life Sciences, Maastricht University, Maastricht; [4]Department of Methodology and Statistics, Medicine and Life Sciences, Maastricht University, Maastricht; [5]School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

Abstract

To develop a tool for use in hearing screening and to evaluate the patient journey towards hearing rehabilitation, responses to *the hearing aid rehabilitation questionnaire* scales *aid stigma*, *pressure*, and *aid unwanted* addressing respectively hearing aid stigma, experienced pressure from others; perceived hearing aid benefit were evaluated with *item response theory*. 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 framework. Items compromising model fit or demonstrating differential item functioning 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. This procedure resulted in bias-free scales ready for screening purposes and application to further understand the help-seeking process of the hearing impaired.

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 behavior 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 that 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 behavior 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 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 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, Vuoriaho 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 40 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 labeled: '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] analyzed 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, 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 audiolo-

gy 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 the 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.

Materials and Methods

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.

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. These scale items are presented in Appendix I (see online Appendices[30-50]).

In Appendix II a detailed description of the methodology applied in this study is given. Classical Test Theory is initially applied to the questionnaire scales, determining whether the assumptions required for IRT modeling, namely uni-dimensionality, local independence and monotonicity are met. Thereafter the IRT models are estimated with assessment of model and item fit and an inspection of generated response and information curves. Differential Item Functioning (DIF) analysis is then performed to ensure that items within each questionnaire scale demonstrate measurement equivalence (are bias free) across groups. This is followed by 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 generated along with the estimated GRM parameters.[34] Model fit at the item level was assessed by examining the $S-\chi^2$ 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 for which it was confirmed that they 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 is considered as statistically significant. A Bonferroni adjustment for multiple com-

parisons 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–χ^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 1 together with the significance levels of the item fit statistics (S–χ^2). In the ICC curves presented in Figure 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 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 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. 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
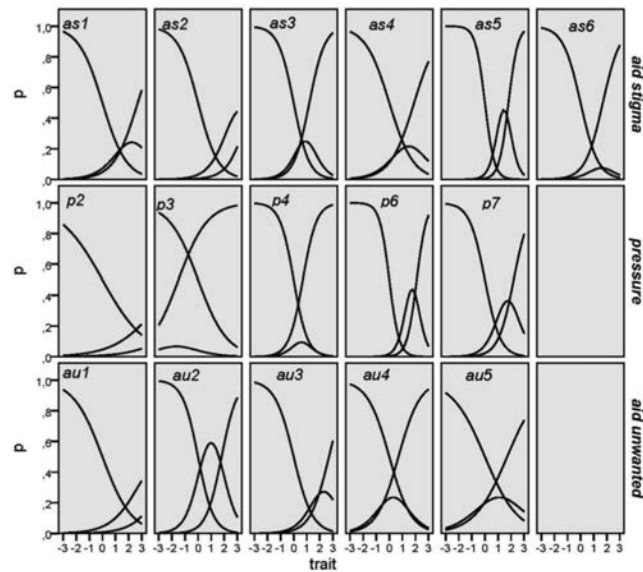

Figure 1. Item characteristic curves of individual items per scale.

Table 1. Initial (full) and final (reduced) scale estimated IRT parameters (se) and item fit statistics of *aid stigma, pressure,* and *aid unwanted* scales with reasons for dropping items.

		Full Scale				Reduced Scale			
	item	a	$b1$	$b2$	S-χ^2	a	$b1$	$b2$	S-χ^2
Aid stigma									
	as1	1.1(0.3)	1.8 (0.4)	2.7 (0.6)	0.62	non uniform DIF for Comparisons I, II			
	as2	1.3 (0.4)	2.5 (0.6)	4.0 (1.1)	0.51	uniform DIF forComparisons I, II			
	as3	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	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	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	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									
	p1	Local independence assumption violated							
	p2	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	1.1 (0.4)	-1.5 (0.4)	-1.2 (0.3)	0.39	uniform DIF Comparison I			
	p4	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								
	p6	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	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									
	au1	0.9 (0.4)	3.2 (1.2)	5.3 (2.1)	0.50	uniform DIF Comparison II			
	au2	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	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	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	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

given in the reduced scale portion at the right in Table 1. Discrimination parameters range from 1.1 to 4.3 with acceptable S–χ^2 statistics for all four items. In the top middle panel of Figure 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 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 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 unaccept-
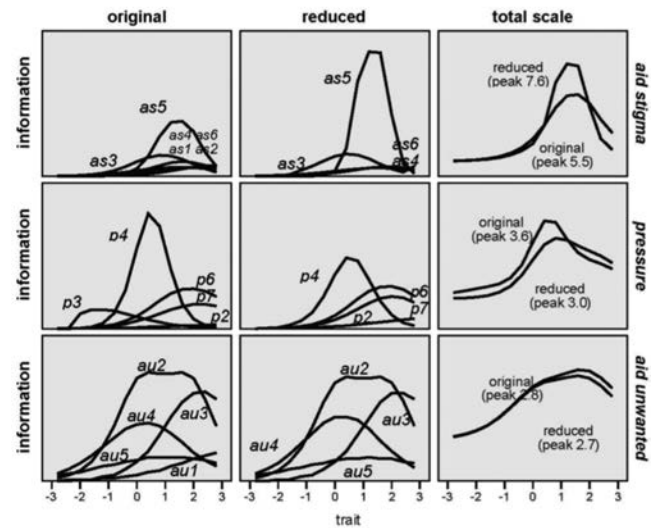

Figure 2. Item information curves of each item relative to other items per scale for the original scale on the left, the same for the reduced scale in the center, and a comparison of total information curves for the original and reduced scales.

Table 2. χ^2-statistics from DIF analysis for Comparisons I: focal = non-hearing impaired; II: focal = hearing aid group; III: focal=hearing aid group, 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 significance at .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 occurring when 0 counts for a response category for one of the groups considered.

able 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–χ^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–χ^2 statistics indicated acceptable item fit. However the estimated discrimination parameter of item *p2* was low at 0.6 as visualized in Figure 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 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–χ^2 item fit statistics. In the center and middle panel of Figure 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 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.

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–χ^2 statistics. In the ICC curves presented in Figure 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 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 modeling of these four remaining items indicated excellent fit statistics and residuals ranging from -0.007 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 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 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
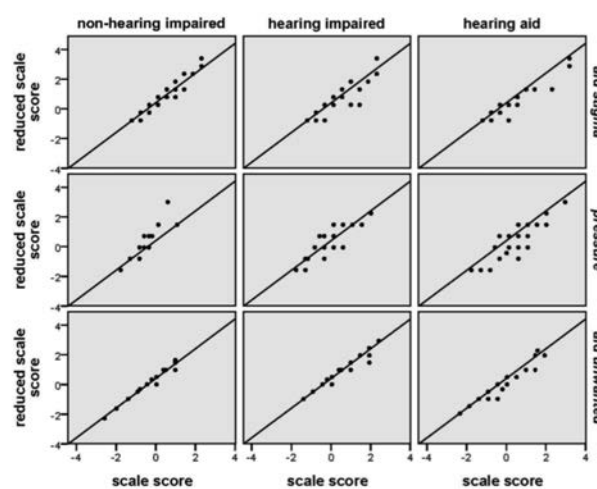

Figure 3. Comparison of scale scores for original and final reduced scales for three groups per scale.

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, both 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–χ^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 redundant items.[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 to 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 determining 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.

Conclusions

In this paper we have examined items in scales addressing factors that may either impede or encourage hearing rehabilitation. The items were from the HARQ questionnaire that was originally developed as a tool to improve hearing rehabilitation primarily for counseling 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 demonstrating measurement equivalence independent of hearing aid use and hearing impairment and also to include items displaying good discriminatory ability along the trait scale for implementation in a hearing screen instrument. A follow-up step is to investigate these refined scales relative 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.

References

1. Davis A, Smith P, Ferguson M, et al. Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models. Health Technol Assess 2007;11:42.

2. Smith PA, Davis AC, Pronk M, et al. Adult hearing screening: what comes next? (Letter to the Editor) Int J Audiol 2011;50:610-2.

3. Manchaiah VKS, Stephens D, Meredith R. The patient journey of adults with hearing impairment: the patients' views. Clin Otolaryngol 2011;36:227-34.

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;8:1265-75.

5. Cox RM, Alexander GC, Gray GA. Who wants a hearing aid? Personality profiles of hearing aid seekers. Ear Hear 2005;26:12-6.

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.

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;30:236-40.

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;30:313-24.

12. Parette P, Sherer M. Assistive technology use and stigma. Educ Train Dev Disabil 2004;39:217-36.

13. Wallhagen M. The stigma of hearing loss. Gerontologist 2009;50:66-74.

14. Duijvestijn JA, Anteunis LJC, Hoek CJ, et al. Help-seeking behaviour of hearing impaired persons aged 55 years, effect of complaints, significant others and hearing aid image. Acta Otolaryngol 2003;123:836-50.

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;28:212-30.

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;41:527-37.

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

18. Joore MA, van der Stel H, Peters HJM, et al. 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, et al. Use of hearing aid and assistive listening devices in an older Australian population. J Am Acad Audiol 2010;21:642-53.

20. Vuorialho A, Sorri M, Nuojua I, Muhli A. Changes in hearing aid use over the past 20 years. Eur Arch Otorhinolaryngol 2006;262:355-60.

21. McCormack A, Fortnum H. Why do people fitted with hearing aids not wear them? Int J Audiol 2013;52:360-8.

22. Cox RM, Alexander GC. Expectations about hearing aids and their relation to fitting outcome. J Am Acad Audiol 2000;14:403-13.

23. Meister H, Walger M, Brehmer D, et al. The relationship between pre-fitting expectations and willingness to use hearing aids. Int J Audiol 2008;47:153-9.

24. Jerram JCK, Purdy SC. Technology, expectations, and adjustment to hearing loss: predictors of hearing aid outcome. J Am Acad Audiol 2001;10:445-57.

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;3:32-41.

26. Hays RD, Lipscomb J. Next steps for use of item response theory in the assessment of health outcomes. Qual Life Res 2007;16:195-9.

27. Demorest ME, Wark DJ, Erdman SA. Development of the screening test for hearing problems. Am J Audiol 2011;20:100-10.

28. Chenault MN, Anteunis LJC, Kremer B, Berger MPF. Quantification of esperienced hearing problems with item resonse theory. Am J Audiol 2013;22:252-62.

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, et al. 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 Suppl 1969;17.

33. Samejima F. Graded response model. In: W.J. van der Linden and R.K. Hambleton (Eds.), Handbook of modern item response theory. Berlin: Springer; 1997. pp 85-100.

34. Chen WH, Thissen D. Local independence indexes for item pairs using item response theory. J Educ Behav Stat 1997;22:265-89.

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;27:289-98.

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. Hillsdale, NJ: Lawrence Erlbaum, Inc.; 1993. pp 67-113.

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: 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;28:283-94.

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;34:120-51.

43. Meister H, Walger M, Brehmer D, et al. The relationship between pre-fitting expectations and willingness to use hearing aids. Int J Audiol 2008;47:153-9.

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;28:181-5.

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;107:238-46.

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;29:168-88.

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; 1985. Available from: http://echo.edres.org:8080/irt/baker/final.pdf