THE GERONTOLOGICAL SOCIETY OF AMERICA®

Innovation in Aging
cite as: *Innovation in Aging*, 2018, Vol. 2, No. 1, 1–11
doi:10.1093/geroni/igy010
Advance Access publication April 12, 2018

OXFORD

Original Research Article

Validity and Reliability of the Major Depression Inventory for Persons With Dual Sensory Loss

Hanna Birkbak Hovaldt, MSc,* Tine Nielsen, PhD, Jesper Dammeyer, PhD

Department of Psychology, Unit of Psychological Cross-Disciplinary and Applied Research, University of Copenhagen, Denmark.

*Address correspondence to Hanna Birkbak Hovaldt, MSc, Department of Psychology, University of Copenhagen, Øster Farimagsgade 2a, 1353 Copenhagen K, Denmark. E-mail: hanna.hovaldt@psy.ku.dk

Received: November 14, 2017; Editorial Decision Date: March 22, 2018

Decision Editor: Laura P. Sands, PhD

Abstract

Background and Objectives: Research has shown that dual sensory loss is a risk factor for depression in older adults. However, validated measures of depression for people with dual sensory loss are lacking. The purpose of the present study was to investigate the construct validity and reliability of the Major Depression Inventory for use among elderly persons with acquired dual sensory loss.

Research Design and Methods: A cross-sectional questionnaire survey was conducted in a national sample of people ≥50 years of age with functional acquired dual sensory loss. Of the invited participants, 302 (66%) returned the questionnaire and 207 complete cases were included for analysis. Rasch models and graphical log-linear Rasch models were used for item analysis. Lack of differential item functioning was tested relative to severity of vision and hearing impairment, mode of questionnaire completion, age, sex, comorbidity, instrumental activities of daily living, social position, and cohabitation status.

Results: The 10-item Major Depression Inventory did not fit the Rasch model. An 8-item version, excluding the items "feeling sad" and "sleep problems," fit a graphical log-linear Rasch model. No evidence of differential item functioning was discovered, thus the 8-item Major Depression Inventory was measurement invariant across severity of impairments and mode of completing the questionnaire. The overall reliability was 0.81 and ranged from acceptable to good for all subgroups of participants, except males with severe hearing impairment and low functional status. Consequently, the 8-item version of the Major Depression Inventory was considered construct valid and reliable within the frame of reference.

Discussion and Implications: An 8-item version of the Major Depression Inventory can be used to screen for depressive symptoms in elderly persons with acquired dual sensory loss.

Translational Significance: An 8-item version of the Major Depression Inventory can be used to validly and reliably screen for symptoms of depression among elderly with acquired dual sensory loss regardless of the severity and duration of their sensory impairments.

Keywords: Depression and anxiety, Disabilities, Psychometrics, Vision

Background and Objectives

Dual sensory loss is a disability where hearing and vision are concurrently impaired, making it impossible to compensate the loss of one sense by use of the other. Dual sensory loss can be either congenital or acquired. Acquired dual sensory loss is largely attributable to age-related

impairments in hearing and vision (Wittich, Watanabe, & Gagné, 2012). Previous research varies regarding the definition of dual sensory loss and the population investigated, but for the majority of persons with dual sensory loss, especially among older adults, the disability is acquired (Wittich et al., 2012). It is reported that the prevalence of dual sensory loss increases from approximately 1% in younger age groups to 1–5% in 65–69 year-olds and to approximately 22% in persons who are 80 years or older, with variations depending on definitions and sampling methods (Schneider et al., 2011). Across studies, it is reported that acquired dual sensory loss is a risk factor affecting communication, cognition, physical functioning, and mental health (Heine & Browning, 2015). Elderly persons with dual sensory loss have consistently been found to have an increased level of symptoms of depression compared with people without sensory loss or people with single sensory loss (Guthrie, Declercq, Finne-Soveri, Fries, & Hirdes, 2016; Heine & Browning, 2014). However, measures of depression can potentially be problematic to use among people with sensory impairments, as some of the symptoms identified by measures of depression could be consequences of the sensory losses rather than depression per se, for example, sleep problems. Sleep problems are a known consequence of visual impairment, due to its impact on the circadian rhythm (Uchiyama & Lockley, 2015). Validity studies of the Patient Health Questionnaire-9 among people with visual impairments have yielded different results in relation to the item on sleep problems (Gothwal, Bagga, & Sumalini, 2014; Lamoureux et al., 2009).

Thus, although there exist potentially problematic items in measures of depression used among older adults with sensory impairments, and these are likely to disrupt accurate estimations of the prevalence of depression in this group, to our knowledge, no thorough examination of the construct validity of measures of depression has so far been conducted in a dual sensory loss population. Only two studies, to our knowledge, report on the internal consistency among persons 18 years or older with dual sensory loss. The two studies investigated the Depression Rating Scale (Burrows, Morris, Simon, Hirdes, & Phillips, 2000) and reported Cronbach's alpha to be 0.68 and 0.63, respectively (Dalby et al., 2009; Guthrie et al., 2011).

Another challenge when surveying people with sensory impairments is communication. People with acquired dual sensory loss can have varying types and degrees of sensory impairments. These differences in the severity and type of sensory impairment can lead to challenges such as the need for a variety of assistive devices and coping strategies in the different subgroups of people with acquired dual sensory loss. As a consequence and to attain a broad representation of people with acquired dual sensory loss, different means of collecting data are necessary to adapt to the sensory functions of the research participants. Examples of such adaptions are tactile language interpretation and having text read aloud either by an assistant or by use of technology. Though these adaptions are needed, they pose a risk of systematically biasing the results of the research performed. To our knowledge, no studies have hitherto investigated whether items in measures of depression function differently among people with acquired dual sensory loss depending on the adaptions in data collection.

The purpose of this study was to investigate the construct validity and reliability of the Major Depression Inventory used among elderly persons with acquired dual sensory loss. Specifically, the focus was on (a) the impact of the degree and duration of vision and hearing impairment on the item responses and (b) the impact of the different adaptions in data collection on the item responses. It was hypothesized that persons with severe vision impairment would systematically be more likely to endorse the item on sleep problems compared to participants with mild vision impairment independent of their score on the Major Depression Inventory. It was further hypothesized that participants who received assistance completing the questionnaire would systematically be less likely to endorse the item "felt life was not worth living" independent of their score on the Major Depression Inventory.

Research Design and Methods

Study Population

The study population consisted of people aged 50 years or older identified with functional acquired dual sensory loss according to the Nordic definition (Ask Larsen & Damen, 2014) by the national provider of services for people with acquired dual sensory loss in Denmark (N = 513). Before data collection, 30 persons were excluded as these persons were not able to complete the survey with assistance due to severe health conditions (e.g., terminal dementia) resulting in a sample size of 483. During the data collection interviewers excluded another 13 persons as they were not able to complete the questionnaire by any means and 14 persons died during the period of data collection. Of the remaining 456 persons, 302 persons (66%) returned the questionnaire and 207 persons (45%) had complete data on the variables included in this study.

Data Collection

The questionnaire was pilot-tested among three persons above 70 years old with acquired dual sensory loss with regard to face validity and the readability of the layout for people with vision loss. Minor changes were made accordingly. The data collection took place from February to and including December 2015. The potential participants were sent an information letter 1–2 weeks before receiving a mailed questionnaire. The information letter explained the project and invited participation by telephone or in-person interviews if completing a paper questionnaire was not feasible. A second questionnaire was sent to nonresponders after 1–2 months. If there was still no response, the

potential participants were contacted by phone and encouraged to participate by the national service provider. They were encouraged to ask for assistance with completion of the questionnaire. Individual contact persons and consultants for the persons with acquired dual sensory loss had been informed about the survey in advance, and they had received an interview guide to enable them to assist in completing the questionnaire if requested. ID numbers served to anonymize the participants. The project was approved by the internal ethics board at University of Copenhagen and by the Danish Data Protection Agency (No. 2015/02).

The Major Depression Inventory

The Major Depression Inventory was developed in 1998 according to the diagnostic criteria for moderate to severe depression in the ICD-10 and major depression in the DSM-IV. The Major Depression Inventory is used to screen for level of depressive symptoms by use of a total scale score and to discriminate between clinical levels of depression by the use of cut-off points (Bech, 2012). The scale score ranges from 0 to 50, and a score of 26 or higher has been suggested as the most appropriate cut-off point for moderate depression (Bech, Rasmussen, Olsen, Noerholm, & Abildgaard, 2001). The Major Depression Inventory differs from the more commonly known Hamilton Depression Rating Scale or Beck Depression Inventory in that it asks about the frequency of depressive symptoms within the last 2 weeks, rather than their intensity (Bech & Wermuth, 1998).

The validity of the Major Depression Inventory has been assessed with regard to dimensionality using classical psychometric methods, nonparametric item response methods, and Rasch models. The inventory has been shown to be unidimensional in general populations (Ellervik, Kvetny, Christensen, Vestergaard, & Bech, 2014) and among patients with depression or Parkinson's disease (Bech & Wermuth, 1998; Konstantinidis, Martiny, Bech, & Kasper, 2011; Olsen, Jensen, Noerholm, Martiny, & Bech, 2003), but not among patients with chronic widespread pain (Amris, Omerovic, Danneskiold-Samsøe, Bliddal, & Wæhrens, 2016). Another study using Mokken analysis found the Major Depression Inventory to be unidimensional only after excluding items 9 and 10 in a sample of people suspected of having depression (Nielsen, Ørnbøl, Vestergaard, Bech, & Christensen, 2017). Furthermore, acceptable sensitivity and specificity (Bech et al., 2001) as well as adequate external validity (Bech, Timmerby, Martiny, Lunde, & Soendergaard, 2015; Olsen et al., 2003) have been established. Item analyses have found less than optimal fit for the items concerning sleep, appetite, and bad conscience (Amris et al., 2016; Nielsen et al., 2017; Olsen et al., 2003), and some items have shown local dependence (Nielsen et al., 2017). In one study using the original 6-point response scale of the Major Depression Inventory, they found that a collapsed 4-point response scale was more appropriate (Nielsen et al., 2017).

Participants in this study were asked to indicate how often during the last 2 weeks each of 10 symptoms had been present, using a 6-point response scale ranging from "all the time" (Score 5) to "none of the time" (Score 0; Table 1). As originally developed, only the highest scores of item 8a and 8b together with 10a and 10b were included in the total score. A low score indicated a low level of depressive symptoms (Bech, 2012). Item 7 originally included the examples "reading newspaper and watching television," but during our pilot-testing it became evident that in the case of low vision the examples were problematic. Thus, "reading newspaper" was substituted with "listening to radio."

Initial analyses showed that some of the response categories were rarely used by our study population. As expected, this caused problems with convergence in the initial analysis. It was therefore decided to collapse the original six response categories into three, and all analyses reported in this study were performed using a 3-point response scale. The choice of cut points was made to achieve meaningful categories, that is, "At no time" (same as original first category), "Less than half the time" (collapse of original second and third categories), and "More than half the time" (collapse of original fourth to sixth categories).

Background Variables Included in Analyses of Differential Item Functioning

All background variables and categories are listed in Table 2. *Hearing impairment* was measured by the question "Do you have a hearing impairment?" with five response categories. Similarly, *Vision impairment* was measured by the question "Do you have a vision impairment?" with five response categories. Slight and moderate impairments were collapsed due to the low number of participants reporting slight impairment and because no-one reported no impairment. *Duration of sensory impairments* was recorded as the number of years since the hearing and vision impairments were first diagnosed by a health professional. *Lawton's instrumental activities of daily living* (Graf, 2008) was used to measure functional status. *Comorbidity* was calculated as the sum of the following conditions, reported to have been diagnosed by a health professional: arthritis (rheumatoid and osteoarthritis), osteoporosis, asthma, lung diseases (chronic bronchitis, emphysema, or chronic obstructive pulmonary disease), coronary thrombosis or angina pectoris, cerebral thrombosis or hemorrhage, high blood pressure, diabetes (Type I or II), stomach or bowel problems, chronic depression, chronic anxiety, prolapsed intervertebral disc or other back problems, cancer, dementia, and a BMI above 30. It was investigated whether instrumental activities of daily living and comorbidity were associated with the item responses independent of depression level, because several of the items could be symptoms or consequences of disease or low physical functioning (Crinion & McNicholas, 2014; Ebede, Jang, & Escalante, 2017). Information about *Level*

Table 1. The Major Depression Inventory and Mean Item Scores (SD) for the 8-item Version ($N = 207$)

	Instruction and items	Mean item scores (SD)	
	The following questions ask about how you have been feeling over the past 2 weeks. Please put a tick in the box which is closest to how you have been feeling. How much of the time…	Original response categories[a]	Collapsed response categories[b]
1.	Have you felt low in spirits or sad?		
2.	Have you lost interest in your daily activities?	1.37 (1.50)	0.88 (0.81)
3.	Have you felt lacking in energy and strength?	2.26 (1.74)	1.25 (0.73)
4.	Have you felt less self-confident?	1.16 (1.43)	0.77 (0.79)
5.	Have you had a bad conscience or feelings of guilt?	0.56 (1.05)	0.39 (0.62)
6.	Have you felt life wasn't worth living?	0.67 (1.24)	0.43 (0.67)
7.	Original: Have you had difficulty in concentrating, e.g., when reading the newspaper or watching television; Our modified version: Have you had difficulty in concentrating, e.g., when listening to the radio or following a television program?	1.32 (1.46)	0.79 (0.85)
8a.	Have you felt very restless?	1.32 (1.46)	0.86 (0.76)
8b.	Have you felt subdued or slowed down?		
9.	Have you had trouble sleeping at night?		
10a.	Have you suffered from reduced appetite?	1.08 (1.60)	0.63 (0.81)
10b.	Have you suffered from increased appetite?		

[a]The six original response categories were: 5 (*all the time*), 4 (*most of the time*), 3 (*slightly more than half the time*), 2 (*slightly less than half the time*), 1 (*some of the time*), and 0 (*at no time*).
[b]The collapsed response categories were: 2 (*all the time, most of the time, or slightly more than half the time*), 1 (*slightly less than half the time, or some of the time*), and 0 (*at no time*).

of education and *Cohabitation status* was obtained. The category "other education" was education/courses that were not convertible to the Danish version of the International Standard Classification of Education. Sociodemographic factors are associated with depression (Alexopoulos, 2005), therefore educational level and cohabitation status as well as age and gender were included for analyses of differential item functioning to ensure that differences in the level of depressive symptoms were not caused by confounding by sociodemographic factors. *Mode of completion* was the method used for completing the survey by the participant and categorized as "No assistance received," "Assistance received from personal relations or professionals," or "Interview by researcher or research assistant." Information about *sex* and *age* was retrieved from the national service provider.

Statistical Analyses

The Rasch models

Construct validity was investigated using the polytomous Rasch Model (Masters, 1982), which is an item response model for ordinal data where the item parameters (i.e., item difficulty) are measured on the same scale as the person parameters (i.e., person ability) of the latent trait (Mesbah & Kreiner, 2013). In this study, an item reflecting a symptom of severe depression is expected to have a high item difficulty and is therefore less likely to be endorsed. Similarly, a person with a high level of depressive symptoms is expectedly more likely to endorse the items of the Major Depression Inventory. A scale fitting the Rasch model has several advantages: (a) the scale is criterion-related construct valid according to the definition of construct validity by Rosenbaum (1989), (b) the scale is specifically objective, meaning that comparisons of persons can be made independent of the items used, and comparisons of items can be made independent of the persons used, (c) the reliability of the scale is as optimal as it can be, and (d) the score of the scale is sufficient for the latent variable, which means that the score includes all the information that can be derived from the items (Kreiner, 2013). To obtain these advantages the data must fit the Rasch model, meaning that they have to meet the following requirements: (a) unidimensionality, all items measure the same construct, (b) monotonicity, the probability of a "high" response to an item increases when the level on the latent variable increases, (c) homogeneity, the rank order of item difficulties are the same for persons with low and high scores on the latent variable, (d) no local dependence of items, meaning that items should not be associated when conditioning on the latent variable, and (e) No differential item functioning (Kreiner, 2013; Mesbah & Kreiner, 2013), meaning that items and relevant exogenous variables should not be associated when conditioning on the latent variable. An example of differential item functioning would include an association between vision impairment severity and item 9 on sleep problems, independent of the Major Depression Inventory score. In this instance, persons with severe vision impairment would have a higher score on the Major Depression Inventory not attributed to higher depression but rather the severity of their vision impairment.

Table 2. Sample Characteristics

	Total (n = 302)		Complete cases (n = 207)	
Characteristic	*N*	%	*N*	%
Sex*				
Male	90	30	54	26
Female	212	70	153	74
Age				
50–64 years	45	15	24	12
65–79 years	58	19	43	21
80–89 years	107	35	77	37
≥90 years	92	30	63	30
Mode of completion*				
No assistance received	34	11	16	8
Assistance received	238	79	175	85
Interview	26	9	16	8
Missing	4	1		
Severity of visual impairment				
Light or moderate vision impairment	45	15	26	13
Severe vision impairment	211	70	155	75
Totally blind	40	13	26	13
Missing	6	2		
Duration of visual impairment				
0–5 years	27	9	21	10
6–10 years	49	16	37	18
11–20 years	72	24	54	26
>20 years	131	43	95	46
Missing	23	8		
Severity of hearing impairment				
Light or moderate hearing impairment	97	32	65	31
Severe hearing impairment	180	60	130	63
Profoundly deaf	19	6	12	6
Missing	6	2		
Duration of hearing impairment				
0–5 years	29	10	24	12
6–10 years	57	19	44	21
11–20 years	69	23	57	28
>20 years	122	40	82	40
Missing	25	8		
Comorbidity*				
0 comorbid conditions	35	12	17	8
≥1 comorbid condition	252	83	190	92
Missing	15	5		
Functional status (IADL score)*				
High functional status (7–8)	43	14	39	19
Medium functional status (4–6)	152	50	127	61
Low functional status (0–3)	59	20	41	20
Missing	48	16		

Note: IADL = instrumental activities of daily living.
*$p < .05$ for difference between complete and incomplete cases (χ^2).

A graphical log-linear Rasch Model is a kind of Rasch model where local dependence and differential item functioning can be adjusted for, if the item responses do not fit the pure Rasch model (Kreiner & Christensen, 2002). If a graphical log-linear Rasch model only includes local dependence, the score will still be sufficient for the latent variable but the reliability will be affected (Hamon & Mesbah, 2002). If a graphical log-linear Rasch model includes differential item functioning, the score of the scale will depend on the level of the exogenous variable causing

differential item functioning and the validity and calculation of the sum score will thus depend on that variable as well (Kreiner & Christensen, 2007).

Item analyses

Initially, fit to the Rasch model was tested. If the item responses did not fit the Rasch model, then analyses of local dependence and differential item functioning were conducted, and subsequent fit to a graphical log-linear Rasch model was tested.

A global test of fit to the model and a global test of differential item functioning were conducted by conditional likelihood ratio tests where item parameters were compared in subpopulations (Andersen, 1973). The fit of the individual items was tested by comparing observed and expected gamma coefficients between items and the rest-scores (Kreiner, 2011). If individual items did not fit the model, they were excluded from the scale. Local dependence, differential item functioning, and associations between theta and background variables were analyzed by partial gamma correlations [γ^p]. γ^p correlations were evaluated as follows: 0–0.1 no correlation, 0.1–0.2 weak correlation, 0.2–0.3 moderate correlation, and >0.3 strong correlation (Nielsen & Kreiner, 2003). Item pairs showing local dependence were combined into composite items and item difficulties were estimated based on these composite items (Kreiner & Christensen, 2007). Differential item functioning was tested relative to severity of vision and hearing impairments, mode of completion, sex, age, duration of vision and hearing impairments, comorbidity, instrumental activities of daily living, highest attained educational level, and cohabitation status. If evidence of differential item functioning was found, item difficulties, reliability, and targeting were stratified by the exogenous variable showing differential item functioning. In case of fit to a graphical log-linear Rasch model, the reliability was estimated using a test–retest simulation with Monte Carlo estimation (Hamon & Mesbah, 2002), and in case of an association between relevant background variables, that is, differences in depression scores for those subgroups, reliability and targeting were estimated for each of these groups. Targeting refers to whether persons (in the study population) are positioned in the interval of the latent variable where there is most information from the items. Targeting was appraised by two indices; the test information target index and the root mean squared error target index. The two indices should be close to one and were calculated as (a) the mean test information divided by the maximum test information and (b) the minimum standard error of measurement divided by the mean standard error of measurement, respectively (Kreiner & Christensen, 2013).

This was a confirmatory validity study and not an inferential analysis of differences in the Major Depression Inventory. It was possible to include persons with missing information on items in graphical log-linear Rasch models, but not persons with missing information on exogenous variables. Only eight participants would be excluded due to missing item responses. The nature of a confirmatory validity study and the fact that the sample would only be increased by eight persons were the two reasons for using complete cases in the analyses. Two subanalyses were decided upon: (a) repeating the analysis excluding instrumental activities of daily living from the models, to include a maximum of cases, while not excluding any crucial information on impairment of hearing or vision, as this was the exogenous variable with most missing values (N = 238), and (b) repeating the analysis with an ordinal version of the education variable, where those with other educations were excluded (N = 188). This was done to test for directional differential item functioning. The results from these subanalyses did not differ substantially from the main analysis and thus only the results from the main analysis are presented in this article.

A critical level of 5% was used in all analyses. Where appropriate, multiple testing was adjusted for using the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995). Descriptive analyses were performed using IBM SPSS 22. Rasch analyses were performed using Digram 3–42 (Kreiner & Nielsen, 2013).

Results

Sample Characteristics

The characteristics of the complete cases (n = 207) included for analyses and the total sample of participants in the questionnaire survey (N = 302) are shown in Table 2. The majority of the complete cases were women (74%) and 80 years or older (67%). For vision impairment, 75% reported having severe vision impairment and 13% reported being totally blind. Of the 207 complete cases, 63% reported having severe hearing impairment and 6% reported being profoundly deaf. Only two of the complete cases reported being totally blind and profoundly deaf. Half of the complete cases (50%) reported having severe vision impairment as well as severe hearing impairment and 20% reported having severe vision impairment and light or moderate hearing impairment. The majority reported having had vision or hearing impairment for more than 10 years (72% and 68%, respectively), and 10% reported having had vision impairment and 12% reported having had hearing impairment for 5 years or less. Of the complete cases, 85% had received assistance completing the questionnaire and 8% had been interviewed either by phone or in person. The majority of the complete cases were living alone (82%). Twenty-two percent of the complete cases had attained a higher education. The remaining cases had vocational (34%), basic (35%), or other (9%) education. Compared to the noncomplete cases, there were significantly more females, more with comorbid conditions, fewer with low functional status, and more who had received assistance completing the questionnaire among the complete cases (Table 2).

Rasch Analyses

Fit analyses

The full 10-item version of the Major Depression Inventory did not fit the Rasch model (Conditional likelihood ratio 34.8, $df = 19$, $p < .05$), and specific evidence of the misfit of items 1 "felt low in spirits or sad" and 9 "trouble sleeping at night" was found ($p < .001$ and $p < .0001$, respectively). Items 1 and 9 were subsequently excluded from the scale. Further analysis by the graphical log-linear Rasch model found evidence of local dependence between the two item pairs; items 2 and 3 ("lost interest in daily activities" and "felt lacking in energy and strength," $\gamma^p = 0.44$, $p < .001$), and items 4 and 5 ("felt less confident" and "had a bad conscience or feelings of guilt," $\gamma^p = 0.50$, $p < .01$; Figure 1). Therefore, two composite items were computed (items 2 + 3 and 4 + 5, respectively). The 8-item version with the two composite items did fit a graphical log-linear Rasch model (results not shown). With regard to differential item functioning, no evidence of differential item functioning by severity or duration of sensory impairments, mode of completion, sex, age, instrumental activities of daily living, comorbidity, educational level, or cohabitation status was found (Table 3). This means that for the two specific focuses, (a) there was no evidence that the Major Depression Inventory functioned differentially relative to the severity and duration of sensory impairments, (b) nor was there any evidence to support the hypothesis that receiving assistance in completing the questionnaire or not would affect the validity of the Major Depression Inventory (i.e., no differential item functioning).

Item difficulty

The item difficulties, meaning the relative difficulty for participants to endorse the different item-statements, are shown in the Supplementary Figure. For all participants, regardless of depression level, the most difficult item to endorse was item 6 "felt life was not worth living." This item required a high depression score to endorse. The easiest item to endorse, regardless of depression level, was the composite item made up of the locally dependent items 2 and 3 "lost interest in daily activities" and "felt lacking in energy and strength."

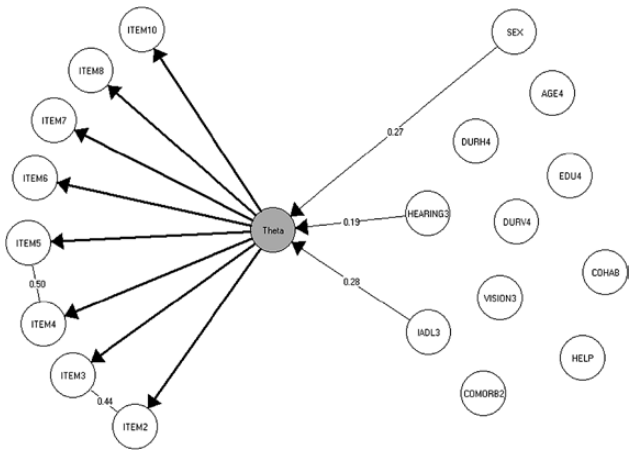

Figure 1. The graphical log-linear Rasch model for the 8-item major depression inventory.

Reliability and targeting

The overall reliability was 0.81. Reliability and targeting were calculated across subgroups defined by gender, severity of hearing impairment and functional status, as these were associated with theta (Figure 1). The reliabilities ranged from 0.58 to 0.87 (Supplementary Table), and males with severe hearing impairment and low functional status had the lowest reliability. Targeting of both theta and the observed score also varied across the subgroups, and was found to be good for all groups except for females with high functional status and any degree of hearing impairment, where only about 63% of the maximum obtainable information was obtained (Supplementary Table).

Mean item scores

The mean score of the 8-item version was 9.81 ($SD = 8.17$) with original scoring (response categories 0–5) and 6.02 ($SD = 4.13$) with modified scoring (response categories 1–3). Item 3 "lacking in energy and strength" had the highest mean item score, whereas item 5 "bad conscience or feelings of guilt" had the lowest (Table 1).

Discussion and Implications

Our results suggest that an 8-item version of the Major Depression Inventory can be used among people with

Table 3. Global Test of Fit (Homogeneity in Score Groups) and Global Tests of Differential Item Functioning of the Graphical Log-Linear Rasch Model for the 8-Item Major Depression Inventory ($N = 207$)

	GLLRM		
Tests of fit	CLR	*df*	*p*
Global homogeneity	11.7	22	.96
Global DIF relative to			
Sex	21.8	22	.47
Age	91.1	66	.02
Educational level	75.1	66	.21
Cohabitation status	34.0	22	.05
Mode of completion	62.6	44	.03
Severity of vision impairment	42.8	44	.52
Duration of vision impairment	76.9	66	.17
Severity of hearing impairment	41.6	22	.58
Duration of hearing impairment	62.1	66	.61
Comorbidity	24.3	22	.33
Functional status	56.7	44	.10

Note: GLLRM includes local dependence interactions between items 2 and 3, and items 4 and 5 respectively. GLLRM = graphical log-linear Rasch model. CLR = conditional likelihood ratio test. DIF = differential item functioning. All *p*-values above .05 after adjustment for false discovery rate due to multiple testing using the Benjamini–Hochberg procedure, as the critical limit was .0042.

acquired dual sensory loss. Furthermore, it seemed like the severity of hearing and vision impairments as well as the mode of data collection did not differentially influence the item responses.

As no other studies, to our knowledge, have used Rasch models to analyze the validity of depression screening instruments among people with acquired dual sensory loss, comparisons with other studies are limited. However, two prior studies have analyzed the use of the Patient Health Questionnaire-9 to measure depression among visually impaired people (Gothwal et al., 2014; Lamoureux et al., 2009) and they both found that the instrument generally had acceptable psychometric properties, which is in line with our findings. One of the studies found that the item "trouble falling or staying asleep" functioned differentially relative to duration of vision impairment, where longer duration was associated with sleep disturbances independent of the level of depression (Gothwal et al., 2014). The other study did not find problems with the sleep item, however, this could be due to lack of statistical power (Lamoureux et al., 2009). In the present study, the item on sleep in the Major Depression Inventory did not fit the Rasch model, this could be due to several reasons. The majority of participants in our study have had vision and/or hearing impairment for more than 10 years. If longer duration of impairments are associated with sleep disturbances, this could be one reason why the item on sleep did not fit the Rasch model. Another explanation could be the effect of visual impairment on the circadian rhythm (Uchiyama & Lockley, 2015). One further explanation could be the high mean age of our sample. Sleep disturbances become more prevalent with increasing age, especially above 60 years of age (Miner & Kryger, 2017). These age-related sleep disturbances can be caused by multiple factors such as morbidity, use of medication, and psychosocial factors (Miner & Kryger, 2017). Future studies should control for factors other than depression to explain sleep problems.

Another item that did not fit the Rasch model in our study was "Have you felt low in spirits or sad?" The item was not found to be directly linked to the sensory impairments, however. In a study among patients suspected of having depression, this item only fit the Rasch model after collapsing the six response categories to four (Nielsen et al., 2017). However, there have been no problems with the fit of the item in other studies (Amris et al., 2016; Olsen et al., 2003). The feeling of being low in spirits or sad is not exclusive to depression, thus, the misfit could be due to participants endorsing the item for other reasons than depression. However, more research is needed in order to determine why this item shows problems with fit to the Rasch model, and whether it might only behave differently among elderly with acquired dual sensory loss.

Altogether, the results propose that two items have to be excluded from the Major Depression Inventory when used among elderly with acquired dual sensory loss. Though the precision of the scale will be lower when omitting the two items, the construct validity will be intact, as the remaining items measure symptoms of depression.

In analyzing differential item functioning, we found no differential item functioning present. Regarding mode of completing the questionnaire, this was in contrast to the hypothesis that participants who received help to complete the questionnaire would respond more positively. However, it should be noted that some of the response categories of mode of completion were not often used, thus this result was not strong. The results showed no evidence of an association between the mode of completing the questionnaire and the observed score on the Major Depression Inventory. In survey research, standardization has been the gold standard for designing survey studies (Bowling, 2005); however, using targeted data collection has been found to reduce nonresponse bias (Lynn, 2017; Rosen et al., 2014). This is in line with our findings which suggest that different modes of data collection in a hard-to-reach group such as elderly persons with acquired dual sensory loss might not bias the findings critically, but rather help to limit nonresponse bias. This should be further examined in future studies.

Reliability of the total scale and the subscales was acceptable except for one subgroup. Reliability levels in this study are in line with those of other studies using the Major Depression Inventory (Bech et al., 2001; Nielsen et al., 2017; Olsen et al., 2003) and among people with sensory impairments (Dalby et al., 2009). The somewhat lower reliabilities found for males might be due to the relatively small number of males in the sample and thus in the subgroups. The targeting was reasonable for the 8-item version of the Major Depression Inventory, which was in line with a previous study on patients suspected of having depression (Nielsen et al., 2017). Depression is strongly associated with functional impairment and disability in elderly populations (Blazer, 2003), which could be the reason why targeting was least optimal for those with high functional status.

Limitations

Though this study's sample is large compared to other dual sensory loss and acquired dual sensory loss studies, it was evident from the distribution of responses on some items of the Major Depression Inventory that some response categories were rarely used. As expected, this caused problems with convergence during initial analysis and response categories were therefore collapsed to three for all analyses reported in this study. Due to this and the exclusion of two items, the scoring of the scale cannot be compared with the original scoring and the original cut-off points can therefore not be used for determining depression among people with acquired dual sensory loss.

Participants were recruited from the national service provider for acquired dual sensory loss in Denmark. Consequently, not all with acquired dual sensory loss in Denmark were identified and included in this study. Thus, the sample might include more individuals with a stronger

network and interest in receiving support causing a potential risk of selection bias. However, the study succeeded in collecting data from a large and well-defined national sample including all individuals identified with acquired dual sensory loss. Due to the support that was provided for data collection, it was possible to include people with severe levels of impairment resulting in a relatively high response rate of 66% in this hard-to-reach population.

This study used self-reported hearing and vision impairment and not objective clinical measures. Among individuals with hearing impairment, it has been reported that clinical measures of hearing impairment are less associated with mental distress than self-reported hearing impairment (Fellinger, Holzinger, Gerich, & Goldberg, 2007). Although self-reported sensory impairment might be the best indicator for mental health, we recommend that future studies include objective clinical measures of sensory impairment. We also recommend that future research ensure complete background information in order to ensure that all relevant subgroups can be included in analyses of differential item functioning. This was not possible for this study, where 207 out of the 302 participants were included for analyses. Lastly, it should be noted that some items such as item 7 on concentration while listening to the radio or following a television program worked well in this sample of individuals with residual vision or hearing, but that this may not be the case for those without residual abilities. Thus, we recommend investigating the validity of the scale among people with complete vision and hearing loss.

Despite the above mentioned limitations, important findings have been discovered due to the use of Rasch models. Modern psychometrics such as Rasch models can identify the potential underlying problems with individual items, and can thus be used to adjust the scale accordingly for use in a specific population of interest. This study is, to our knowledge, the first study to use Rasch models in an acquired dual sensory loss sample underlining the need for the application of modern psychometrics to measures used among elderly persons with severe health conditions. Future research should retest the 8-item Major Depression Inventory and investigate its convergent validity, sensitivity, and specificity.

Conclusion

This study found an 8-item version of the Major Depression Inventory to be appropriate as a screening instrument for depressive symptoms among elderly persons with acquired dual sensory loss. Interestingly, and perhaps surprisingly, we found that the severity of hearing and vision losses might not differentially influence the scoring of the Major Depression Inventory (no differential item functioning). This indicates that symptoms associated with depression and acquired dual sensory loss can possibly be measured independently, and the scale can possibly be used among people with mild as well as severe sensory impairments. These results underline the significance of validating instruments before use in populations of elderly persons with severe health conditions.

Finally, the different modes of completion did not seem to lead to response bias in this study. However, as some of the modes of completion were seldom used, this result requires further examination in future research. Future studies among elderly persons with severe health conditions affecting their ability to participate in survey studies should investigate whether the studies might benefit from a multimodal data collection approach to increase the response rate and reduce nonresponse bias.

Supplementary Material

Supplementary data are available at *Innovation in Aging* online.

Funding

This work was supported by the Velux Foundation (VELUX33847).

Acknowledgements

The researchers would like to thank the participants with acquired dual sensory loss and CFD for access to data.

Conflict of Interest

None.

References

Alexopoulos, G. S. (2005). Depression in the elderly. *Lancet (London, England)*, **365**, 1961–1970. doi:10.1016/S0140-6736(05)66665-2

Amris, K., Omerovic, E., Danneskiold-Samsøe, B., Bliddal, H., & Wæhrens, E. E. (2016). The validity of self-rating depression scales in patients with chronic widespread pain: A Rasch analysis of the major depression inventory. *Scandinavian Journal of Rheumatology*, **45**, 1–11. doi:10.3109/03009742.2015.1067712

Andersen, E. B. (1973). A goodness of fit test for the rasch model. *Psychometrika*, **38**, 123–140. doi:10.1007/BF02291180

Ask Larsen, F., & Damen, S. (2014). Definitions of deafblindness and congenital deafblindness. *Research in Developmental Disabilities*, **35**, 2568–2576. doi:10.1016/j.ridd.2014.05.029

Bech, P. (2012). *Clinical Psychometrics*. Oxford, UK: Wiley-Blackwell. doi:10.1002/9781118511800

Bech, P., Rasmussen, N. A., Olsen, L. R., Noerholm, V., & Abildgaard, W. (2001). The sensitivity and specificity of the major depression inventory, using the present state examination as the index of diagnostic validity. *Journal of Affective Disorders*, **66**, 159–164. doi:10.1016/S0165-0327(00)00309-8

Bech, P., Timmerby, N., Martiny, K., Lunde, M., & Soendergaard, S. (2015). Psychometric evaluation of the major depression inventory (MDI) as depression severity scale using the LEAD (longitudinal expert assessment of all data) as index of validity. *BMC Psychiatry*, **15**, 190. doi:10.1186/s12888-015-0529-3

Bech, P., & Wermuth, L. (1998). Applicability and validity of the Major Depression Inventory in patients with Parkinson's disease. *Nordic Journal of Psychiatry*, **52**, 305–310. doi:10.1080/08039489850149741

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, **57**, 289–300. doi:10.2307/2346101

Blazer, D. G. (2003). Depression in late life: Review and commentary. *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, **58**, 249–265. doi:10.1093/gerona/58.3.M249

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health (Oxford, England)*, **27**, 281–291. doi:10.1093/pubmed/fdi031

Burrows, A. B., Morris, J. N., Simon, S. E., Hirdes, J. P., & Phillips, C. (2000). Development of a minimum data set-based depression rating scale for use in nursing homes. *Age and Ageing*, **29**, 165–172. doi:10.1093/ageing/29.2.165

Crinion, S. J., & McNicholas, W. T. (2014). Sleep-related disorders in chronic obstructive pulmonary disease. *Expert Review of Respiratory Medicine*, **8**, 79–88. doi:10.1586/17476348.2014.860357

Dalby, D. M., Hirdes, J. P., Stolee, P., Strong, J. G., Poss, J., Tjam, E. Y.,…Ashworth, M. (2009). Development and psychometric properties of a standardized assessment for adults who are deafblind. *Journal of Visual Impairment & Blindness*, **103**, 7–16.

Ebede, C. C., Jang, Y., & Escalante, C. P. (2017). Cancer-related fatigue in cancer survivorship. *The Medical Clinics of North America*, **101**, 1085–1097. doi:10.1016/j.mcna.2017.06.007

Ellervik, C., Kvetny, J., Christensen, K. S., Vestergaard, M., & Bech, P. (2014). Prevalence of depression, quality of life and antidepressant treatment in the Danish General Suburban Population Study. *Nordic Journal of Psychiatry*, **68**, 507–512. doi:10.3109/08039488.2013.877074

Fellinger, J., Holzinger, D., Gerich, J., & Goldberg, D. (2007). Mental distress and quality of life in the hard of hearing. *Acta Psychiatrica Scandinavica*, **115**, 243–245. doi:10.1111/j.1600-0447.2006.00976.x

Gothwal, V. K., Bagga, D. K., & Sumalini, R. (2014). Rasch validation of the PHQ-9 in people with visual impairment in South India. *Journal of Affective Disorders*, **167**, 171–177. doi:10.1016/j.jad.2014.06.019

Graf, C. (2008). The Lawton instrumental activities of daily living scale. *The American Journal of Nursing*, **108**, 52–62. doi:10.1097/01.NAJ.0000314810.46029.74

Guthrie, D. M., Declercq, A., Finne-Soveri, H., Fries, B. E., & Hirdes, J. P. (2016). The health and well-being of older adults with dual sensory impairment (DSI) in four countries. *PLoS One*, **11**, e0155073. doi:10.1371/journal.pone.0155073

Guthrie, D. M., Pitman, R., Stolee, P., Strong, G., Poss, J., Tjam, E. Y.,…Hirdes, J. P. (2011). Reliability of standardized assessment for adults who are deafblind. *Journal of Rehabilitation Research and Development*, **48**, 545–554. doi:10.1682/JRRD.2010.09.0175

Hamon, A., & Mesbah, M. (2002). Questionnaire reliability under the Rasch model. In M. Mesbah, B. F. Cole, & M. T. Lee (Eds.), *Statistical methods for quality of life studies, design, measurements, and analysis* (pp. 155–168). Dordrecht, Holland: Kluwer Academic Publishers. doi:10.1007/978-1-4757-3625-0_13

Heine, C., & Browning, C. J. (2014). Mental health and dual sensory loss in older adults: A systematic review. *Frontiers in Aging Neuroscience*, **6**, 1–9. doi:10.3389/fnagi.2014.00083

Heine, C., & Browning, C. (2015). Dual sensory loss in older adults: a systematic review. *The Gerontologist*, **55**, 913–928. doi:10.1093/geront/gnv074

Konstantinidis, A., Martiny, K., Bech, P., & Kasper, S. (2011). A comparison of the major depression inventory (MDI) and the beck depression inventory (BDI) in severely depressed patients. *International Journal of Psychiatry in Clinical Practice*, **15**, 56–61. doi:10.3109/13651501.2010.507870

Kreiner, S. (2011). A note on item–restscore association in Rasch models. *Applied Psychological Measurement*, **35**, 557–561. doi:10.1177/0146621611410227

Kreiner, S. (2013). The Rasch model for dichotomous items. In K. B. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch models in health* (pp. 5–25). London: Wiley. doi:10.1002/9781118574454.ch1

Kreiner, S., & Christensen, K. B. (2002). Graphical Rasch models. In M. Mesbah, B. F. Cole, & M. T. Lee (Eds.), *Statistical measures for quality of life studies* (pp. 187–203). Dordrecht, Holland: Kluwer Academic Publishers. doi:10.1007/978-1-4757-3625-0_15

Kreiner, S., & Christensen, K. B. (2007). Validity and objectivity in health-related scales: Analysis by graphical loglinear Rasch models. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution rasch models: Extensions and applications* (pp. 329–346). New York: Springer. doi:10.1007/978-0-387-49839-3_21

Kreiner, S., & Christensen, K. B. (2013). Person parameters estimation and measurement in Rasch models. In K. B. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch models in health* (pp. 63–78). London: Wiley. doi:10.1002/9781118574454.ch4

Kreiner, S., & Nielsen, T. (2013). *Item analysis in DIGRAM 3.04. Part I: Guided tours. Research report 2013/06*. Copenhagen: Department of Public Health, University of Copenhagen. https://ifsv.sund.ku.dk/biostat/annualreport/images/0/01/Research_Report_13-06-ny.pdf

Lamoureux, E. L., Tee, H. W., Pesudovs, K., Pallant, J. F., Keeffe, J. E., & Rees, G. (2009). Can clinicians use the PHQ-9 to assess depression in people with vision loss? *Optometry and Vision Science*, **86**, 139–145. doi:10.1097/OPX.0b013e318194eb47

Lynn, P. (2017). From standardised to targeted survey procedures for tackling non-response and attrition. *Survey Research Methods*, **11**, 93–103. doi:10.18148/srm/2017.v11i1.6734

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, **47**, 149–174. doi:10.1007/BF02296272

Mesbah, M., & Kreiner, S. (2013). Rasch models for ordered polytomous items. In K. B. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch models in health* (pp. 27–41). London: Wiley. doi:10.1002/9781118574454.ch2

Miner, B., & Kryger, M. H. (2017). Sleep in the aging population. *Sleep Medicine Clinics*, **12**, 31–38. doi:10.1016/j.jsmc.2016.10.008

Nielsen, T., & Kreiner, S. (2003). Tests til afprøvning af hypoteser inddragende to variable. In: *SPSS - Introduktion til databehandling & statistisk analyse* (2nd ed., pp. 265–288). Denmark: Jurist- og Økonomiforbundets Forlag.

Nielsen, M. G., Ørnbøl, E., Vestergaard, M., Bech, P., & Christensen, K. S. (2017). The construct validity of the major depression inventory: A Rasch analysis of a self-rating scale in primary care. *Journal of Psychosomatic Research*, **97**, 70–81. doi:10.1016/j.jpsychores.2017.04.001

Olsen, L. R., Jensen, D. V., Noerholm, V., Martiny, K., & Bech, P. (2003). The internal and external validity of the major depression inventory in measuring severity of depressive states. *Psychological Medicine*, **33**, 351–356. doi:10.1017/S0033291702006724

Rosen, J. A., Murphy, J., Peytchev, A., Holder, T., Dever, J., Herget, D., & Pratt, D. (2014). Prioritizing low propensity sample members in a survey: Implications for nonresponse bias. *Survey Practice*, **7**, 1–10. doi:10.29115/SP-2014-0001

Rosenbaum, P. R. (1989). Criterion-related construct validity. *Psychometrika*, **54**, 625–633. doi:10.1007/BF02296400

Schneider, J. M., Gopinath, B., McMahon, C. M., Leeder, S. R., Mitchell, P., & Wang, J. J. (2011). Dual sensory impairment in older age. *Journal of Aging and Health*, **23**, 1309–1324. doi:10.1177/0898264311408418

Uchiyama, M., & Lockley, S. W. (2015). Non-24-hour sleep-wake rhythm disorder in sighted and blind patients. *Sleep Medicine Clinics*, **10**, 495–516. doi:10.1016/j.jsmc.2015.07.006

Wittich, W., Watanabe, D. H., & Gagné, J. P. (2012). Sensory and demographic characteristics of deaf-blindness rehabilitation clients in Montréal, Canada. *Ophthalmic & Physiological Optics*, **32**, 242–251. doi:10.1111/j.1475-1313.2012.00897.x