Calibration of an item bank in 474 orthopedic patients using Rasch analysis for computer-adaptive assessment of anxiety

Clinical Rehabilitation
2019, Vol. 33(9) 1468–1478
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0269215519846225
journals.sagepub.com/home/cre

Selina Kallinger¹ , Henry Scharm¹, Maren Boecker²,
Thomas Forkmann³ and Harald Baumeister¹

Abstract

Objective: To calibrate an item bank of anxiety-related questions for use in orthopedic patients within a computer-adaptive test.

Design: This is a psychometric study.

Setting: The sample of orthopedic patients was recruited in two orthopedic rehabilitation clinics in Germany.

Subjects: A total of 474 orthopedic rehabilitation patients were recruited for this study.

Interventions: Not applicable.

Main measures: The main measure is an adapted version of an existing anxiety item pool for cardiovascular rehabilitation patients.

Results: The results of the confirmatory factor analysis and Mokken analysis confirmed a one-factor structure and double monotonicity. An anxiety item bank (48 items) could be developed and calibrated using Rasch analysis. It fitted to the Rasch model with a non-significant item–trait interaction ($\chi^2(203) = 172.59$; $P = .94$) and was free of differential item functioning. Unidimensionality could be verified and the person separation reliability was .96. The category threshold parameters varied between 4.72 and 3.16 (7.88 logits).

Conclusion: The unidimensional anxiety item bank provides the basis for a computer-adaptive test to assess a wide range of anxiety in rehabilitation patients with orthopedic diseases with very good psychometric characteristics.

Keywords

Anxiety, orthopedic disease, computer-adaptive test, item bank, item response theory, Rasch model

Received: 5 July 2018; accepted: 28 March 2019

¹Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education, University of Ulm, Ulm, Germany

²Institute of Medical Psychology and Medical Sociology, University Hospital of RWTH Aachen, Aachen, Germany

³Department of Clinical Psychology, University of Duisburg-Essen, Essen, Germany

Corresponding author:

Selina Kallinger, Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education, University of Ulm, Albert-Einstein-Allee 47, D-89081 Ulm, Germany.

Email: selina.kallinger@uni-ulm.de

Introduction

Anxiety disorders are the most frequent mental disorders in orthopedic rehabilitation patients next to affective disorders with a prevalence of 15%.¹ Orthopedic patients are seemingly more often affected with an anxiety disorder than healthy persons and patients with other somatic diseases.^{2–5} In orthopedic rehabilitation patients, the presence of anxiety and depressive symptoms is associated with symptom aggravation of the physical illness, health impairment, lower probability of remission and more than 50% less response to treatment.^{4,5} To counteract these negative consequences, it is highly relevant to screen for anxiety in orthopedic patients at early stage.

In rehabilitation clinics, the detection rate of mental disorders is insufficient. In orthopedic patients, approximately, only 50% of mental disorders are detected.^{6–8} Therefore, an initial and follow-up assessment of anxiety should be integrated in daily rehabilitation routine.^{7,9} This demand conflicts with limited psychodiagnostic resources like less psychological staff and restricted time.^{7,10}

A computer-based assessment provides an opportunity to save time and personnel resources. Especially, computer-adaptive tests (CAT) might counteract the aforementioned challenges.^{10–16} To this end, an unidimensional item bank is required. To our knowledge, none of the existing item banks for the assessment of anxiety is focusing especially on orthopedic rehabilitation patients.^{17–19} This seems as more important as somatic anxiety items in medically ill patients are regarded as highly problematic due to potential symptom overlap, which might be the case in generic anxiety item banks not tested for patients with somatic diseases.^{19,20}

This study aimed at checking whether the AIB-cardio could also be applied for diagnosing severity of anxiety in orthopedic patients.¹⁹ In case no model fit was found for AIB-cardio, a disease-specific item bank excluding somatic aspects of anxiety should be developed and calibrated to overcome the lack of a missing item bank for the assessment of anxiety, specifically in orthopedic patients (AIB-ortho). The development and calibration included testing the test fairness with regard to orthopedic diseases and using a homogeneous item and response format.

Methods

The sample of orthopedic patients was recruited in two German orthopedic rehabilitation clinics. Based on the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10),²¹ chapter XIII: Diseases of the musculoskeletal system and connective tissue (M00-M99), we included patients with arthrosis (ICD-10: M15-M19),²¹ spondylopathies (ICD-10: M45-M49), other dorsopathies (ICD-10: M50-M54), other soft tissue disorders (ICD-10: M70-M79), and other ICD-10-M-diagnoses or other orthopedic diagnosis except M-diagnosis. Exclusion criteria were insufficient German language skills, dementia or acute intoxication. Ethical committee approval was granted and study procedures were in accordance with ethical principles stated in the Declaration of Helsinki.²²

Development and calibration of the anxiety item bank was part of the project “Implementation of computer-adaptive initial- and follow-up assessment of health-related functioning in orthopedic and cardiological rehabilitation” funded by the German statutory pension insurance scheme (*Deutsche Rentenversicherung Bund*). The aim of this project is to develop and implement a computer-adaptive software (RehaCAT) to assess the health-related functioning of cardiovascular and orthopedic patients with the domains “Anxiety,” “Depression,” “Functioning in daily living,” “Functioning of upper limbs,” “Functioning of lower limbs” and “work capacity.” The two participating orthopedic rehabilitation clinics implemented RehaCAT in their daily routine. The anxiety item pool was presented on a computer as an in-depth assessment for patients meeting the inclusions criteria. Patients had to self-rate the whole item pool on the computer. Skipping of items was not possible.

The basis of the developed item pool was the final item pool of the AIB-cardio¹⁹ and consisted of 155 items. First, nine items were eliminated because of irrelevance for orthopedic patients (e.g. ... have you been worried about suffering of a stroke?) and eight items were reformulated with a special focus on orthopedic illnesses. Second, four items considering fear of progression were added.^{23,24} The introductory

phrase was consistent across all remaining 150 items (“How often during the last two weeks ...?”). The time interval of two weeks should ensure that state anxiety was measured. The response format was a unified 5-point Likert-type scale (0=never; 1=rarely; 2=sometimes; 3=mostly; 4=always). All items are available on request from the corresponding author.

Medical record data were used to describe the sample. Socio-demographics (gender, age, nationality, family status, educational level, employment) were assessed via self-report.

Data analysis

The basis of each CAT is a calibrated unidimensional item bank.¹⁴ Unidimensional means that the responses to all items are determined by one latent construct (in this study: anxiety). The term “calibrated” indicates assigned item difficulty levels and standard errors to all unidimensional items.^{19,25–27}

A CAT is based on a computer algorithm, which creates an individual and optimal test for each patient.¹² The item selection process of the algorithm starts with an item consisting of mean difficulty and selects the next items based on the previous responses until a predefined stopping rule is reached.^{13–15} This process leads to a substantial item reduction in 50% to 90% without loss of measurement precision.¹¹ CATs are flexible and highly test fair.^{15,16} In addition, a CAT allows user-oriented interfaces and automatic evaluations of the patient records, for example, in written or graphic reports.

A three-step analysis strategy was used starting with confirmatory factor analysis (CFA), followed by Mokken analysis and finally Rasch analysis.

CFA. A CFA was conducted using Mplus²⁸ to verify the single-factor structure. The weighted least square estimator (WLSM/WLSMV) was used due to the skewed categorical data. To examine model fit, the root mean square error of approximation (RMSEA < .06: good fit), standardized root mean residuals (SRMR < .08: good fit), comparative fit index (CFI > .95: good fit) and Tucker–Lewis index (TLI > .95: good fit) were used.^{29,30} The factor loading cut-off was fixed at communalities < .3.

Mokken analysis. A Mokken³¹ analysis was conducted using R to confirm the assumption of the double monotonicity as a requirement of the partial credit Rasch model (PCM). The assumption of the double monotonicity contains: (1) all patients have the same hierarchical order to the probability to solve the item; (2) all items have the same hierarchical order to the probability to get solved by the patients. To check this assumption, the coefficient H of Loevinger (weak scale: $.3 \leq H < .4$; medium scale: $.4 \leq H < .5$; strong scale: $.5 \leq H$) was used.³²

Rasch analysis. The Rasch model is a model of the item response theory (IRT) and enables framework of CATs and development of new scales.^{27,33–35} The Rasch model is based on the assumption that the person’s response to an item only depends on that person’s characteristic on the measured dimension (e.g. degree of anxiety) and on the difficulty of the item.³⁶ The probability of a person’s response is a logistic function of the relative distance between item localization and person localization on a linear scale.³⁷ The item analysis was conducted using RUMM2030.³⁸ Due to a significant likelihood ratio test ($P < .001$), the partial credit model (PCM) was preferred over the rating scale model.^{37,39,40} The approach of the Rasch analysis contained the following iterative steps and critical values.^{19,37}

One important step is threshold ordering. A threshold is the transition point of one response category to another where the probability of acceptance of both response categories is the same. Considering the used 5-point Likert-type scale, there are four thresholds per item. If a threshold is disordered (e.g. the probability of response category 1 is smaller than category 0), the threshold should be ordered. Ordering thresholds means to combine the “problematic”—adjacent thresholds (in the previous example, response category 1 should be rescored to 0). For this purpose, the category probability curves were inspected.^{19,37,41}

To evaluate the fit to the model, these parameters were used: item–trait interaction score for the overall fit statistic (χ^2 , $P > .05$) and individual item fit (fit residuals values $> \pm 2.5$; item- χ^2 -values ($P < .05$, Bonferroni-adjusted)).^{19,37,40} Items with misfit were excluded during the Rasch analysis process.

The items of an item bank should be independent of each other as local response dependency (LD) inflates reliability and introduces bias in parameter estimation. If LD was found, one item of the pair was initially excluded to develop an LD-free core item bank. A cut-off value of $+ .20$ above the average observed residual correlation was used.^{42,43} Later on, we tried to re-enter locally dependent items using anchoring (see below). The cut-off value was calculated for every conducted analysis.

Differential item functioning (DIF) is an unexpected interaction with sample characteristics, caused by subgroups (e.g. gender) who answer differently to an item, although identical underlying level of anxiety.^{19,27,37,44} The analyses of DIF were calculated for gender, age and previously mentioned specific orthopedic diagnoses. DIF analyses were calculated by an analysis of variance of the person-item-deviation residuals with person factors and class intervals as factors.^{19,27,37,45} Items with DIF were excluded.

To verify the unidimensionality of the final item bank, the approach of Smith⁴⁶ was used. In RUMM2030, two subsets of items (positive vs. negative loadings of the first residual factor) were generated to perform independent *t* tests. An item bank is considered unidimensional if the 95% confidence interval of the number of significant tests overlaps 5%.^{19,37,46,47}

The targeting of the scale was visually illustrated by the person-item threshold distribution graph. The person-item distribution compares the location of items (fixed to zero logits) with the location of persons.^{19,37,47}

The reliability of the item bank was checked using the person separation index (PSI). For group testing, a PSI value of $.7$ to $.85$ is considered as acceptable, whereas for individual testing, a PSI value above $.85$ should be achieved.³⁷

After developing a misfit-, DIF-, and LD-free core item bank, items which were excluded because of LD were tried to be re-added by anchoring to enlarge the item bank. First, an anchor set consisting of non-LD items and covering a wide range of difficulties was prepared. Threshold estimates of the development of the core item bank were used as anchors. Using this anchor set, we tried item by

item, whether the initially excluded items would fit into the item bank by testing all mentioned critical values. If fit to the Rasch model was achieved, the item was re-added to the item bank. Of course, in later CAT applications, LD items should never be presented to the same person.

Results

Finally, 474 patients were recruited (for more details, see Table 1).

Results of CFA

Seven items of the initial item pool (150 items) were deleted due to low factor loadings (Figure 1). The remaining 143 items had factor loadings between $.50$ and $.87$. The model fit confirmed an unidimensional solution, as the RMSEA was smaller than $.06$ (RMSEA = $.04$), the SRMR was below $.08$ (SRMR = $.06$) and the CFI and TLI were close to $.95$ (CFI = $.92$; TLI = $.92$).

Results of Mokken analysis

A total of 21 items were deleted due to low values of Loevinger's *H* (Figure 1). The remaining item pool (122 items) confirmed double monotonicity as a prerequisite of the PCM (Loevinger's *H* = $.52$).

Results of Rasch analysis

The AIB-cardio (37 items)¹⁹ did not fit to the Rasch model showing a significant probability value of the item-trait interaction score ($\chi^2(259) = 464.43$; $P < .001$). A total of 17 residual correlations were above the critical cut-point and seven items with misfits could be identified. Therefore, a disease-specific orthopedic anxiety item bank had to be developed.

A total of 51 items were deleted because of individual-item misfit, 7 items were deleted because of DIF (-3 items $\rightarrow$ DIF diagnoses; -4 items $\rightarrow$ DIF gender) and 35 were deleted because of LD (Figure 1). In all, 5 items of the remaining 29 items had to be rescored (3 had to be merged into 3 categories and 2 had to be merged into 4 categories) (Table 2).

Table 1. Characteristics of the sample (N=474).

	N	%
Gender		
Female	346	73.0
Male	128	27.0
Age in years ($\bar{X}$ 52.5 $\pm$ 7.9)		
$\leq$ 51	179	37.8
52–57	154	32.5
$\geq$ 58	141	29.7
Marital status		
Single	67	14.1
Married/ in partnership	324	68.4
Divorced/ separated	69	14.6
Widowed	14	3.0
Educational level		
Up to 9 years of school ^a	119	25.1
Up to 10 years of school	230	48.5
Over 10 years of school	125	26.4
Employment status		
Employed	426	89.9
Unemployed	45	9.5
No specification	3	0.6
Specific orthopedic ICD-10 diagnoses		
M15-19: arthrosis	44	9.3
M45-49: spondylopathies	15	3.2
M50-54: other dorsopathies	318	67.1
M70-79: other soft tissue disorders	40	8.4
Other M-diagnoses	36	7.6
Other orthopedic diagnoses (except M-diagnoses)	21	4.4
Nationality		
German	463	97.7
Other	11	2.3

ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10th Revision.

^aNo graduation (n=4; 0.8%); no specification (n=5; 1.1%).

The core item bank (29 items) showed fit to the Rasch model with a non-significant probability value of the item–trait interaction score (χ^2 (203)=172.59; P =.94). Statistics of the residuals for items ($\bar{x}$ =-.20; SD=1.12) and persons ($\bar{x}$ =-.30; SD=1.23) supported good model fit compared to the perfect values ($\bar{x}$ =0; SD=1). Item fit residual values (-2.26 to 1.89) varied inbetween the uncritical range ($\pm$ 2.5). Item- χ^2 -values varied between 2.6 and 11.75 with a probability value range of .11–.92.

All 29 items of the core item bank were free of DIF and LD. Unidimensionality of the item bank was supported by the results of the paired t tests (24 of 468 t tests (5.1%) were significant) as the 95% confidence interval (1.9%; 8.3%) overlaps the required 5%.

The category threshold parameters of the item bank covered a range of 7.88 logits (-4.72 to 3.16) and captured a wide spectrum of anxiety (Figure 2). The location of the items varied between -1.40 and 2.65 logits ($\bar{x}$ =.00; SD=.93) (Table 2).

The mean location of persons was -1.74 logits (SD=1.69) with six extreme persons (1.27%). This indicates that the sample had a lower level of anxiety than the average level of anxiety measured by the item bank. The PSI was .96.

A total of 19 items (of potential 35 items with LD) were re-added to the item bank by anchoring (Figure 1). Three items had to be rescored (two had to be merged into three categories and one had to be merged into four categories (Table 2)). Item fit residual values (-1.4 to 2.01) varied between the uncritical range ($\pm$ 2.5). Item- χ^2 -values varied between 4.51 and 17.59 with a probability value range of .01 to .59 (all items had a probability value above the Bonferroni-adjusted alpha value).

All 19 items of the anchored items were free of DIF. Due to anchoring, there were 33 LD items specified (Table 2). The location of the items varied between -1.09 and .49 logits (Table 2).

Discussion

The overall aim of this study was to develop and calibrate an item bank for orthopedic rehabilitation patients to assess anxiety. We were successful generating such an item bank containing 48 items with good fit indices that could serve as the groundwork for an CAT measuring a wide range of anxiety in this special patient group.

Because the item pool (150 items) of this study almost consisted of the same item pool items as AIB-cardio,¹⁹ we first checked whether the AIB-cardio could also be used for orthopedic patients. However, the Rasch analysis of the AIB-cardio¹⁹ demonstrates that this item bank does not fit to orthopedic patients. Therefore, it was necessary to

Figure 1. Flowchart of item selection.

calibrate an item bank specifically for orthopedic patients. This might be indicative for the general need for disease-specific item banks at least for anxiety or, respectively, for thorough psychometric evaluations of generic anxiety item banks such as the item bank of Walter et al. and Pilkonis et al. in populations with different medical conditions.^{17,18}

The obtained AIB-ortho (48 items) shows very good properties: the CFA and Mokken analysis indicated a one-factor solution and the calibration by Rasch analysis showed very good fit to the PCM.³⁹ Although the CFI and TLI could not reach the required cut-off criteria, unidimensionality could be assumed considering all results. There is evidence that RMSEA is more robust to effects of number of items than CFI and TLI.⁴⁸ The item bank covers a wide spectrum of anxiety and measures cognitive and emotional aspects of anxiety. The AIB-ortho enables to assess very severe anxiety precisely due to many difficult items. Only eight items had to be rescored. This supports the selected 5-point Likert-type scale and shows that the patients were able to discriminate between the response options. In regard to a disease-specific

item bank, it is important to mention that the AIB-ortho is free of DIF regarding to gender, age and orthopedic diagnoses. Hence, the assessment based on AIB-ortho is independent of orthopedic-specific disease, and in addition, the AIB-ortho can be used optimally for individual or group testing.³⁷ Due to anchoring, it was possible to enlarge the core item bank (29 items) to 48 items (AIB-ortho). In CAT applications, it should be ensured that LD items are never presented to the same person. The number of items is fully sufficient for using it to develop specific assessment tools such as CATs^{27,47,49–51} and anxiety short forms.^{52,53} A CAT adapts questions based on previously given answers. Therefore, the person's ability has to be estimated anew after every item and the next one has to be selected by an information criterion. To calculate all these accurately, information about answers from a large representative sample are required. The AIB-ortho provides such a psychometrically sound basis for the development of a CAT for assessing anxiety in orthopedic rehabilitation patients. Given the consistent item and response format, it might minimize the considerable burden

Table 2. Rasch parameters of the anxiety item bank for orthopedic patients (AIB-ortho).

How often during the last two weeks		Location (SE)	Residual (df)	Chi-square (probability)	Scoring structure	LD
Core item bank						
AIB-O-01	... have you been thinking you are crazy?	2.65 (0.18)	-1.23 (448)	7.87 (0.34)	00112	
AIB-O-02	... have you been afraid to die soon?	2.01 (0.14)	-0.41 (448)	7.76 (0.35)	00112	
AIB-O-03	... have you had a blank in your mind?	1.36 (0.10)	-1.63 (448)	6.16 (0.52)	00123	✓
AIB-O-04	... have you been worried about being unable to express your opinion?	1.20 (0.12)	1.25 (448)	5.11 (0.65)	00122	✓
AIB-O-05	... have you been afraid of rather everyday activities?	0.96 (0.07)	-0.50 (448)	5.93 (0.55)	01234	✓
AIB-O-06	... have you experienced difficulties controlling your emotions?	0.74 (0.07)	-1.74 (448)	5.87 (0.55)	01234	
AIB-O-07	... have you felt more anxious than normal?	0.32 (0.07)	-1.09 (448)	6.74 (0.46)	01234	✓
AIB-O-08	... have you been afraid of losing control?	0.27 (0.07)	-0.65 (448)	5.56 (0.59)	01234	
AIB-O-09	... have unpleasant thoughts come unintentionally to your mind?	0.21 (0.07)	0.11 (448)	3.40 (0.85)	01234	✓
AIB-O-10	... have you been afraid of not being able to cope with trivial tasks?	0.16 (0.07)	0.44 (448)	6.61 (0.47)	01234	
AIB-O-11	... have you been afraid that something terrible will happen?	0.12 (0.07)	0.45 (448)	11.45 (0.12)	01234	
AIB-O-12	... have you been confused?	0.01 (0.07)	0.41 (448)	6.41 (0.49)	01234	✓
AIB-O-13	... have you been afraid of becoming helpless?	-0.08 (0.06)	0.29 (448)	2.60 (0.92)	01234	✓
AIB-O-14	... have you been afraid to show your feelings?	-0.16 (0.06)	0.31 (448)	4.25 (0.75)	01234	✓
AIB-O-15	... have you been worried when noticing an unpleasant feeling in your body?	-0.17 (0.07)	0.89 (448)	3.57 (0.83)	01234	✓
AIB-O-16	... have you had the feeling that your thinking was slower than usual?	-0.24 (0.06)	1.73 (448)	5.37 (0.62)	01234	✓
AIB-O-17	... have thoughts apparently been running through your head?	-0.38 (0.06)	-1.30 (448)	6.08 (0.53)	01234	✓
AIB-O-18	... have you had a lack of self-confidence?	-0.40 (0.07)	-0.59 (448)	2.60 (0.92)	01234	✓
AIB-O-19	... have you wondered about your failures and shortcomings?	-0.41 (0.06)	0.08 (448)	3.44 (0.84)	01234	✓
AIB-O-20	... have you taken everything seriously?	-0.50 (0.07)	-2.26 (448)	11.75 (0.11)	01234	
AIB-O-21	... have you been agitated?	-0.51 (0.07)	-1.12 (448)	5.77 (0.57)	01234	
AIB-O-22	... have you felt tense?	-0.56 (0.06)	0.34 (448)	6.29 (0.51)	01234	
AIB-O-23	... have you had a bad feeling?	-0.64 (0.07)	-1.13 (448)	6.80 (0.45)	01234	
AIB-O-24	... have you been concerned that something could go wrong?	-0.64 (0.07)	-0.06 (448)	5.74 (0.57)	01234	
AIB-O-25	... have you believed that your problems became too much for you?	-0.65 (0.07)	-1.59 (448)	6.36 (0.50)	01233	✓
AIB-O-26	... have you been nervous?	-0.97 (0.07)	1.75 (448)	4.14 (0.76)	01234	
AIB-O-27	... have you had disturbing thoughts go through your mind?	-1.16 (0.07)	-1.14 (448)	5.17 (0.64)	01234	
AIB-O-28	... have you been annoyed?	-1.17 (0.07)	1.89 (448)	7.76 (0.35)	01234	✓
AIB-O-29	... have you been afraid?	-1.40 (0.07)	0.71 (448)	6.03 (0.54)	01234	✓

(Continued)

Table 2. (Continued)

How often during the last two weeks		Location (SE)	Residual (df)	Chi-square (probability)	Scoring structure	LD
LD items						
AIB-O-30	... have you been worried about not taking care for yourself?	0.49 (0.10)	1.29 (357)	11.27 (0.13)	00122	✓
AIB-O-31	... have you had less feelings than usual?	0.41 (0.07)	-0.34 (355)	14.18 (0.05)	01234	✓
AIB-O-32	... have you wondered about managing life worse as other people?	0.36 (0.07)	0.87 (357)	11.25 (0.13)	01234	✓
AIB-O-33	... have you been frightened?	0.30 (0.07)	-0.10 (357)	10.06 (0.19)	01234	✓
AIB-O-34	... have you had more anxiety and worries as other people would have in your situation?	0.30 (0.07)	-0.73 (355)	9.99 (0.19)	01234	✓
AIB-O-35	... have you been afraid of losing your job?	0.29 (0.09)	1.61 (357)	9.66 (0.21)	00112	✓
AIB-O-36	... have you been worried about not getting accepted by others?	0.25 (0.07)	1.77 (360)	13.79 (0.06)	01234	✓
AIB-O-37	... have you been easily embarrassed?	0.24 (0.07)	0.82 (356)	17.59 (0.01)	01234	✓
AIB-O-38	... have you been hesitant?	0.23 (0.07)	-0.18 (362)	4.51 (0.72)	01234	✓
AIB-O-39	... have you had difficulties thinking clearly?	0.16 (0.07)	-0.12 (345)	4.77 (0.69)	01234	✓
AIB-O-40	... have you been incapable of controlling your thoughts?	0.13 (0.07)	-0.67 (343)	7.05 (0.42)	01234	✓
AIB-O-41	... have you been worried about making a fool of yourself?	0.12 (0.07)	2.00 (362)	9.18 (0.24)	01234	✓
AIB-O-42	... have you had difficulties thinking of other things than your health?	0.09 (0.06)	1.47 (357)	13.10 (0.07)	01234	✓
AIB-O-43	... have you felt guilty?	-0.01 (0.06)	1.51 (358)	9.38 (0.23)	01234	✓
AIB-O-44	... have your worries and fears interfered with everyday activities?	-0.12 (0.07)	-1.41 (358)	8.53 (0.29)	01234	✓
AIB-O-45	... have you been worried about losing contact to close friends?	-0.14 (0.07)	0.76 (356)	10.71 (0.15)	01233	✓
AIB-O-46	... have you been somewhere else with your thoughts?	-0.32 (0.07)	1.47 (369)	4.92 (0.67)	01234	✓
AIB-O-47	... have you been sorrowed?	-0.41 (0.07)	-0.26 (366)	11.87 (0.11)	01234	✓
AIB-O-48	... have you been incapable to relax?	-1.09 (0.06)	2.01 (372)	12.19 (0.09)	01234	✓

SE: standard error; df: degrees of freedom; LD: local dependency; location: location in logits.

Figure 2. Person-item threshold distribution of the AIB-ortho.

for patients by a short processing time.¹⁹ Nevertheless, the AIB-ortho can also be used for the development of short-form questionnaires to allow for anxiety testing where computers are not available.^{52,53} Moreover, AIB-ortho builds a further puzzle for a comprehensive psychodiagnostics in orthopedic and cardiovascular patients using RehaCAT with its already established IRT and CAT dimensions depression, activity in daily living and work capacity as well as the parallel to this study developed item bank on patients' treatment motivation.^{19,27,45,47,49–54,55}

This study shows one limitation: The potential for exact measurement of a very mild anxiety is limited. This limitation, however, seems to be negligible, as clinicians working with patients are most likely not interested in very mild to absent anxiety states.

In conclusion, the newly developed and calibrated item bank (AIB-ortho) shows very good psychometric properties and provides the basis for a CAT in orthopedic rehabilitation patients. The AIB-ortho covers a wide range of anxiety. Of particular note in this context is the potential of AIB-ortho to assess severe anxiety. Thereby, the comprehensive developing process and the statistical procedure

indicate a very high construct validity of the AIB-ortho. The next step will be to develop a CAT based on the AIB-ortho as well as parallel short forms and to examine the convergent and discriminant validity of these tests. With regard to the development of the AnxietyCAT_{ortho}, clinical practice tests are needed to determine the efficiency of the AIB-ortho as the basis for a CAT.

Clinical messages

- Anxiety in orthopedic rehabilitation patients should be assessed by specific rehabilitation instruments as the developed anxiety item bank (AIB-ortho).
- The AIB-ortho seems to be a valid instrument that is sensitive across a wide spectrum of anxiety.

Acknowledgements

The authors are grateful to all patients who participated in this study. They would like to specifically thank Reha-Zentrum Bad Driburg (Klinik Berlin) and Reha-Zentrum Bad Pyrmont (Klinik Weser) for support of data collection.

Author contributions

S.K. performed the study, analyzed the data and wrote the paper; H.S. performed the study; M.B. conceived and designed the study, analyzed the data; T.F. conceived and designed the study; H.B. conceived, designed and performed the study. All authors contributed to interpretation of the data and writing up of the manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the German Insurance Scheme (Grant No. 8011-106-31/31.117). It was conducted within the framework of the project “Implementation of computer-adaptive initial- and follow-up assessment of health-related functioning in orthopedic and cardiological rehabilitation.” The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

ORCID iD

Selina Kallinger <https://orcid.org/0000-0002-2209-0516>

References

- Härter M, Hahn D, Baumeister H, et al. Komorbidität bei Patienten mit muskuloskelettalen und kardiovaskulären Erkrankungen. *J Public Health* 2004; 12: 162–167.
- De Moraes VY, Jorge MR, Faloppa F, et al. Anxiety and depression in Brazilian orthopaedics inpatients: a cross sectional study with a clinical sample comparison. *J Clin Psychol Med Settings* 2010; 17(1): 31–37.
- Härter M, Baumeister H, Reuter K, et al. Increased 12-month prevalence rates of mental disorders in patients with chronic somatic diseases. *Psychother Psychosom* 2007; 76(6): 354–360.
- Baumeister H and Härter M. Psychische Komorbidität bei muskuloskelettalen Erkrankungen. *Bundesgesundheitsblatt* 2011; 54: 52–58.
- Matcham F, Norton S, Scott DL, et al. Symptoms of depression and anxiety predict treatment response and long-term physical health outcomes in rheumatoid arthritis: secondary analysis of a randomized controlled trial. *Rheumatology* 2016; 55(2): 268–278.
- Härter M, Woll S, Reuter K, et al. Recognition of psychiatric disorders in musculoskeletal and cardiovascular rehabilitation patients. *Arch Phys Med Rehabil* 2004; 85(7): 1192–1197.
- Jahed J, Bitz I, Bengel J, et al. Dokumentation komorbider psychischer Störungen in der medizinischen Rehabilitation: Eine Analyse ärztlicher Entlassungsberichte. *Rehabil* 2012; 51: 385–391.
- Reuter K, Woll S, Stadelmann S, et al. Erkennen und Behandeln psychischer Belastungen und Störungen in der orthopädischen Rehabilitation. *Z Klin Psychol Psychiatr Psychother* 2002; 3: 313–327.
- Reuter K and Härter M. Diagnostik psychischer Belastungen und Störungen bei körperlichen Erkrankungen. In: Härter M, Baumeister H and Bengel J (eds) *Psychische Störungen bei körperlichen Erkrankungen*. Heidelberg: Springer, 2007, pp.15–28.
- Elhan AH, Öztuna D, Kutlay S, et al. An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. *BMC Musculoskelet Disord* 2008; 9: 166.
- Gibbons RD, Weiss DJ, Kupfer DJ, et al. Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatr Serv* 2008; 59(4): 361–368.
- Meijer RR and Nering ML. Computerized adaptive testing: overview and introduction. *Appl Psychol Meas* 1999; 23: 187–194.
- Ware JE, Sinclair SJ, Gandek B, et al. Item response theory and computerized adaptive testing. *Rehabil Psychol* 2005; 50: 71–78.
- Van der Linden WJ. *Computerized adaptive testing: theory and practice*. Boston, MA: Kluwer Academic, 2000.
- Reeve BB. Special issues for building computerized adaptive tests for measuring patient-reported outcomes. *Med Care* 2006; 44(11 suppl 3): S198–S204.
- Gershon RC. Computer adaptive testing. *J Appl Meas* 2005; 6: 109–127.
- Pilkonis PA, Choi SW, Reise SP, et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. *Assessment* 2011; 18: 263–283.
- Walter OB, Becker J, Björner JB, et al. Development and evaluation of a computer adaptive test for “Anxiety” (Anxiety-CAT). *Qual Life Res* 2007; 16(suppl 1): 143–155.
- Abberger B, Haschke A, Krense C, et al. The calibrated, unidimensional anxiety item bank for cardiovascular patients provided the basis for anxiety assessment in cardiovascular rehabilitation patients. *J Clin Epidemiol* 2013; 66(8): 919–927.
- Zigmond AS and Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983; 67: 361–370.
- World Health Organization (WHO). *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10)*. Geneva: WHO, 2010.
- World Medical Association. World Medical Association Declaration of Helsinki—ethical principles for medical

- research involving human subjects. *J Am Med Assoc* 2013; 310: 2191–2194.
23. Berg P, Book K, Dinkel A, et al. Progredienzangst bei chronischen Erkrankungen. *Psychother Psych Med* 2011; 61: 32–37.
 24. Kwakkenbos L, van Lankveld WG, Vonk MC, et al. Disease-related and psychosocial factors associated with depressive symptoms in patients with systemic sclerosis, including fear of progression and appearance self-esteem. *J Psychosom Res* 2012; 72(3): 199–204.
 25. Hambleton RK. Emergence of item response modeling in instrument development and data analysis. *Med Care* 2000; 38(9 suppl): II60–II65.
 26. Millman J and Arter JA. Issues in item banking. *J Educ Meas* 1984; 21: 315–330.
 27. Forkmann T, Böecker M, Norra C, et al. Development of an item bank for the assessment of depression in persons with mental illnesses and physical diseases using Rasch analysis. *Rehabil Psychol* 2009; 54(2): 186–197.
 28. Muthén LK and Muthén BO. *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
 29. Hu LT and Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999; 6: 1–55.
 30. Browne MW and Cudeck R. Alternative ways of assessing model fit. *Sociol Methods Res* 1992; 21: 230–258.
 31. Mokken RJ. *A theory and procedure of scale analysis*. Berlin: Walter de Gruyter, 1971.
 32. Mokken RJ. Nonparametric models for dichotomous responses. In: Linden WJ and Hambleton RK (eds) *Handbook of modern item response theory*. New York: Springer, 1997, pp.350–367.
 33. Embretson SE and Reise SP. *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
 34. Rasch G. *Probabilistic models for some intelligence and attainment tests*. Oxford: Nielsen & Lydiche, 1960.
 35. Gilworth G, Chamberlain MA, Bhakta B, et al. Development of the BD-QoL: a quality of life measure specific to Behçet's disease. *J Rheumatol* 2004; 31: 931–937.
 36. Hays RD, Morales LS and Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care* 2000; 38(9 suppl): II28–II42.
 37. Tennant A and Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it. *Arthritis Rheum* 2007; 57(8): 1358–1362.
 38. Andrich D, Lyne A, Sheridan B, et al. RUMM2030, <http://www.rummlab.com.au/>
 39. Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982; 47: 149–174.
 40. Lamoureux EL, Pallant JF, Pesudovs K, et al. The Impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci* 2006; 47(11): 4732–4741.
 41. Pallant JF, Miller RL and Tennant A. Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis. *BMC Psych* 2006; 6: 28.
 42. Christensen KB, Makransky G and Horton M. Critical values for Yen's Q3: identification of local dependence in the Rasch model using residual correlations. *Appl Psychol Meas* 2017; 41(3): 178–194.
 43. Marais I. Local dependence. In: Christensen KB, Kreiner S and Mesbah M (eds) *Rasch models in health*. London: Wiley-Blackwell, 2013, pp.111–130.
 44. Tennant A and Pallant JF. DIF matters: a practical approach to test if differential item functioning makes a difference. *Rasch Meas Trans* 2007; 20: 1082–1084.
 45. Haschke A, Abberger B, Müller E, et al. Calibration of an item bank for work capacity in cardiological rehabilitation patients. *Eur J Prev Cardiol* 2013; 20(5): 807–816.
 46. Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002; 3: 205–231.
 47. Baumeister H, Abberger B, Haschke A, et al. Development and calibration of an item bank for the assessment of activities of daily living in cardiovascular patients using Rasch analysis. *Health Qual Life Outcomes* 2013; 11: 133.
 48. Cook KF, Kallen MA and Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res* 2009; 18: 447–460.
 49. Abberger B, Haschke A, Wirtz M, et al. Development and evaluation of a computer adaptive test to assess anxiety in cardiovascular rehabilitation patients. *Arch Phys Med Rehabil* 2013; 94(12): 2433–2439.
 50. Forkmann T, Kroehne U, Wirtz M, et al. Adaptive screening for depression—recalibration of an item bank for the assessment of depression in persons with mental and somatic diseases and evaluation in a simulated computer-adaptive test environment. *J Psychosom Res* 2013; 75(5): 437–443.
 51. Scholz M, Müller E, Kröhne U, et al. Validierung der computeraadaptiven Assessmentprozedur RehaCAT hinsichtlich funktionaler und psychischer Belastungen bei Rehabilitanden mit muskuloskelettalen Erkrankungen. *Rehabil* 2014; 53: 290–296.
 52. Abberger B, Haschke A, Tully PJ, et al. Development and validation of parallel short forms PaSA-cardio for the assessment of general anxiety in cardiovascular rehabilitation patients using Rasch analysis. *Clin Rehabil* 2017; 31(1): 104–114.
 53. Schmucker A, Abberger B, Boecker M, et al. Parallel short forms for the assessment of activities of daily living in cardiovascular rehabilitation patients (PADL-cardio): development and validation. *Disab Rehabil* 2017; 41: 826–832.
 54. Haschke A, Abberger B, Schröder K, et al. Überprüfung kalibrierter Itembanken zur Erfassung der beruflichen Funktionsfähigkeit an einer Stichprobe ambulanter kardiologischer Rehabilitanden. *Rehabil* 2013; 52: 383–390.
 55. Müller E, Bengel J and Wirtz M. Validität und Änderungssensitivität der Skalen zur Erfassung der Funktionsfähigkeit im Beruf (SE-FFB). *Rehabil* 2013; 53: 176–183.

This text is made available via DuEPublico, the institutional repository of the University of Duisburg-Essen. This version may eventually differ from another version distributed by a commercial publisher.

DOI: 10.1177/0269215519846225

URN: urn:nbn:de:hbz:464-20210505-122138-2

This publication is with permission of the rights owner freely accessible due to an Alliance licence and a national licence (funded by the DFG, German Research Foundation) respectively.

© The Author(s) 2019. All rights reserved.