TLAR
1947

Original Article

Psychometric Properties of the ICF Core Set for Low Back Pain and Its Clinical Use

ICF Bel Ağrısı Çekirdek Setinin Psikometrik Özellikleri ve Klinik Kullanımı

Derya ÖZTUNA,[1] Burcu YANIK,[3] Şehim KUTLAY,[2] Yeşim KURTAİŞ AYTÜR,[2] Atilla Halil ELHAN,[1]
Alan TENNANT,[4] Ayşe Adile KÜÇÜKDEVECİ[2]

Departments of [1]Biostatistics, [2]Physical Medicine and Rehabilitation, Medical Faculty of Ankara University, Ankara, Turkey;
[3]Department of Physical Medicine and Rehabilitation, Medical Faculty of Fatih University, Ankara, Turkey;
[4]Department of Rehabilitation Medicine, Faculty of Medicine and Health, University of Leeds, UK

Objectives: In this study, we investigated the psychometric properties of the International Classification of Functioning, Disability and Health (ICF) core set for low back pain (LBP).

Patients and methods: One-hundred outpatients with LBP (73 females, 27 males; mean age 55.3 years; range 24 to 84 years) were assessed by the ICF core set for LBP. The patients also completed the Roland-Morris disability questionnaire (RMDQ) and Short Form-36 (SF-36) questionnaire. The internal construct validity of the ICF core set for LBP was assessed by Rasch analysis and external construct validity by correlations with the RMDQ and SF-36 Health Survey version 1.0. Reliability was tested by internal consistency and person separation index.

Results: After rescoring the disordered response categories and deletion of some items, "body functions and body structures" and "activities and participation" item sets satistifed Rasch model expectations with a mean item fit of 0.005 (SD 0.619) and −0.006 (SD 0.730), and person fit of −0.165 (SD 0.561) and −0.084 (SD 0.806) respectively. Both item sets were unidimensional and showed no differential item functioning. Their reliabilities were good with Cronbach's alpha coefficient and person separation index levels above 0.77. Although the mean functionality level of the patients was lower than the mean difficulty level of the items, the distribution of the difficulty level of the items overlapped with the distribution of the functionality level of the patients for both item sets. The presence of the expected level of correlations between both item sets and RMDQ and SF-36 has confirmed the external construct validity. "Environmental factors" did not meet the assumptions of the Rasch analysis.

Conclusion: After some modifications, a 15-item "body functions and body structures" set and a 21-item "activities and participation" set from the ICF comprehensive core set were found to be reliable and valid to assess functioning in patients with LBP.

Key words: Disability; low back pain; Rasch analysis; validity and reliability.

Amaç: Bu çalışmada "İşlevsellik, Yetiyitimi ve Sağlığın Uluslararası Sınıflandırması" (ICF) bel ağrısı çekirdek setinin psikometrik özellikleri incelendi.

Hastalar ve yöntemler: Ayaktan tedavi olan bel ağrılı 100 hasta (73 kadın, 27 erkek; ort. yaş 55.3 yıl, dağılım 24-84 yıl), bel ağrılı hastalar için belirlenen ICF çekirdek seti ile değerlendirildi. Hastalar Roland-Morris disabilite anketi (RMDQ) ve Kısa Form-36 (KF-36) anketini de doldurdu. Bel ağrısına yönelik ICF çekirdek setinin iç yapısal geçerliliği Rasch analizi ile dış yapısal geçerliliği ise RMDQ ve KF-36 Sağlık Araştırması versiyon 1.0 ile ilişkileri incelenerek değerlendirildi. Güvenirlik, içsel tutarlılık ve birey ayırsama indeksi kullanılarak test edildi.

Bulgular: Sıra takip etmeyen yanıt kategorileri yeniden skorlandıktan ve bazı maddeler çıkarıldıktan sonra, "vücut fonksiyonları ve vücut yapıları" ve "aktiviteler ve katılım" madde setleri Rasch modeli beklentilerini karşıladı ve ortalama madde uyumu sırasıyla 0.005 (SS 0.619) ve −0.006 (SS 0.730), ortalama birey uyumu ise sırasıyla −0.165 (SS 0.561) ve −0.084 (SS 0.806) bulundu. Her iki madde setinin de tek boyutlu olduğu, madde işlev farklılığı göstermediği belirlendi. Her iki setin de güvenilirlikleri yüksek olup, Cronbach'ın alfa katsayısı ve birey ayırsama indeksi değerleri 0.77'nin üzerindeydi. Her bir madde seti için, hastaların işlevsellik düzeyi ortalaması, maddelerin zorluk düzeyi ortalamasından düşük olsa da, maddelerin zorluk düzeylerinin dağılımı ile hastaların işlevsellik düzeylerinin dağılımı örtüştü. Her iki madde setinin de, RMDQ ve KF-36 ile beklenen düzeyde korelasyon göstermesi, dış yapısal geçerliliklerini göstermekte idi. "Çevresel faktörler" bileşeni Rasch analizinin varsayımlarını sağlamadı.

Sonuç: Gerekli düzenlemeler sonrasında, 15 maddelik "vücut fonksiyonları ve vücut yapıları" seti ve 21 maddelik "aktiviteler ve katılım" seti, bel ağrılı hastalarda işlevselliğin değerlendirilmesinde güvenilir ve geçerli olarak bulundu.

Anahtar sözcükler: Disabilite; bel ağrısı; Rasch analizi; geçerlik ve güvenirlik.

Low back pain (LBP) is a frequent musculoskeletal problem causing disability.[1] The assessment of disability is essential for both planning and monitoring therapeutic interventions in the routine clinical management of patients with LBP. There are many scales available for outcome assessment in LBP, most of which measure impairment and activity limitation. Only a few of these include participation in society.[2,3]

The International Classification of Functioning, Disability and Health (ICF) developed by the World Health Organization (WHO) aims to provide a unified and standard language and framework for the description of health and health-related conditions. It describes a model which systematically classifies the health and health-related domains into two components: (i) body functions (BF) and body structures (BS) and (ii) activities and participation (AP). According to this model, functioning is an umbrella term encompassing all BF, BS and AP while disability is an umbrella term including both impairments and activity limitations or participation restriction. Body functions refer to physiological functions of body systems whereas BS are anatomical parts of the body. Impairments are problems in body function or structure such as a significant deviation or loss. Activity is the execution of a task or action by an individual and represents the individual perspective of functioning. Participation is involvement in a life situation and represents the societal perspective of functioning. The ICF also lists environmental factors (EF) that interact with all these constructs.[4]

The ICF classification comprises 1545 categories divided into four components (BF, BS, AP, EF). In order to make this comprehensive classification applicable in health care, ICF Core Sets which are short lists of ICF categories relevant for specific conditions were developed. Currently, there are ICF core sets for various musculoskeletal conditions including LBP.[5]

The description of functioning based on the ICF involves the rating of ICF categories with the ICF qualifiers. These are numeric codes that specify the extent or the magnitude of functioning in that category or the extent to which an EF is a facilitator or barrier. Qualifier ratings across a number of ICF categories result in an ordinal profile. An ordinal profile may provide a useful tool for healthcare interventions. The important question is whether it is possible to use this profile as a measurement instrument for an ICF component. Thus, the aim of this study was to investigate the reliability and construct validity of the ICF Comprehensive Core Set for LBP as a potential assessment tool for functioning. To accomplish this aim, the reliability and construct validity of components of this ICF Core Set were tested by both modern and classical psychometric methods.

PATIENTS AND METHODS

Patients and setting

Data was collected in the Department of Physical Medicine and Rehabilitation at the Medical School of Ankara University, Turkey. A total of 100 outpatients (73 females, 27 males; mean age 55.3±16.7 years; range 24 to 84 years) with LBP were included in the study. Patients with non-mechanical back pain resulting from inflammatory, infectious, malignant or visceral diseases or with a history of recent surgery that could affect assessment were excluded. The Ethical Committee of Ankara University approved the study and all patients gave written informed consent.

Assessment

The assessment included the administration of the ICF Core Set for LBP, the Roland-Morris disability questionnaire (RMDQ) for LBP[6] and the Short Form-36 Health Survey version 1.0 (SF-36®).[7] The scoring of the ICF Core Set for all patients was performed by rehabilitation medicine specialists who were trained in a structured one-day workshop organized by the researchers of the WHO ICF Collaborating center at the Ludwig-Maximilian University in Munich. The questionnaires RMDQ and SF-36 were either self-completed by literate patients or administered by assessors to illiterates. Sociodemographic (age, gender, years of education, employment status) and clinical data (disease duration, etiology, disease severity) were also recorded.

The ICF Core Set for LBP consists of 78 ICF categories organized in four different components of which BF contains 19 categories, BS five categories, AP 29, and EF 25 categories. A generic qualifier scale was used to evaluate the extent of a patient's problem in each of the ICF categories. The qualifier scale of the components BF, BS and AP has five response levels ranging from 0 to 4: no/mild/moderate/severe/complete problem. The qualifier scale of the component EF has nine response levels ranging from −4 to +4. A specific EF can be a barrier (−1 to −4), or a facilitator (+1 to +4), or can have no influence (0) on a patient's life. If a factor has an influence, the extent of the influence (either positive or negative) can be coded as mild, moderate, severe, or complete. For the Rasch

analysis, scoring of EF items was done as 0 for −4, 1 for −3, 2 for −2, 3 for −1, 4 for 0, 5 for +1, 6 for +2, 7 for +3 and 8 for +4. In addition, there are the response options "8 (not specified)" and "9 (not applicable)" for all ICF categories of all components.[4] In our analysis, "8 (not specified)" and "9 (not applicable)" responses were accepted as missing values.

Physical disability due to LBP was assessed by the RMDQ. It includes 24 items, each with a dichotomous response category of yes or no. The scale has a total score ranging from 0 to 24 with a high score showing higher disability. The Turkish version of the RMDQ was used.[7]

The health-related quality of life was evaluated using the SF-36 questionnaire.[8] It contains 36 items that measure perceived health in eight scales (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health) with higher scores (range 0-100) reflecting better perceived health. Additionally, two summary scores can be obtained- the physical component summary score and the mental component summary score. The Turkish version of the SF-36 was used in the study.[9]

Internal construct validity

The internal construct validity of each component of the ICF Core Set for LBP '"BF and BS", AP and EF items' was assessed by Rasch analysis. Rasch analysis is the formal testing of an assessment or a scale against a mathematical measurement model which defines how interval scale measurements can be derived from ordinal questionnaires.[10-12] The Rasch model assumes that the probability of a given respondent affirming an item is a logistic function of the difference between the item difficulty and the person ability parameter. Master's partial credit model (PCM) which is an extension of the Rasch dichotomous model for polytomous (more than two response categories) items was used in this study.[13]

Common fundamental attributes of the Rasch model were assessed.[14] These are (i) the appropriate stochastic ordering of response categories; (ii) fit of items and persons to the model; (iii) test of the assumption of the local independence of items, including response dependency and unidimensionality; and (iv) the presence of differential item functioning (DIF).

As one of the most common sources of item misfit concerns respondents' inconsistent use of these response categories, the response categories should be examined for correct ordering of thresholds before the evaluation of item fit where polytomous items are involved. For an item with an appropriate ordering of thresholds, thresholds should increase in their location in a manner consistent with the increase in the underlying trait being measured. When this does not occur, the thresholds are said to be disordered, and the categories may have to be collapsed to ensure that this is the case.[15]

A range of fit statistics is used to test if the data conform to Rasch model expectations. Two are item-person interaction statistics transformed to approximate a Z score representing a standardized normal distribution. If the items and persons fit the model, we would expect to see a mean of approximately zero and a standard deviation (SD) of one. The third is a summed chi-square within groups defined by their position on the trait where the overall chi-square for items is summed to give the item-trait interaction statistic. This tests the property of invariance across the trait. A significant chi-square indicates that the hierarchical ordering of the items varies across the trait which compromises the required property of invariance. In addition to these overall summary fit statistics, individual person- and item-fit statistics are presented as (i) residuals (a summation of individual person and item deviations) and (ii) as a chi-square statistic. Fit residuals between ±2.5 are deemed to be adequate. These are summated within ability groups to provide the basis of the analysis of variance (ANOVA).[14,15]

A formal test of the assumption of unidimensionality is undertaken by performing a principal component analysis (PCA) of the residuals. Items with the highest positive and negative correlations on the first residual PC are used to construct two smaller scales that are anchored to the item difficulties of the main analysis.[16] The person estimates derived from these two subsets of items are contrasted for each individual by a t-test. A significant difference would be expected to occur by chance in 5% of the cases. Consequently, the percentage of t statistic outside the range ±1.96 is reported together with a 95% binomial confidence interval. This interval should overlap 5% for a non-significant finding to confirm unidimensionality.

The assumption of local independence implies that when the 'Rasch factor' has been extracted, there should be no leftover patterns in the residuals.

Performing a PCA analysis of the residuals obtained from PCM tested this assumption. If a pair of items had a residual correlation of 0.30 or more, one of the items that showed a higher accumulated residual correlation with the remaining items was eliminated.[17]

Items are also tested for DIF. In the framework of Rasch measurement, the scale should be free of item bias or DIF.[18] Differential item functioning occurs when different groups within the sample (e.g., younger and older persons) respond in a different manner to an individual item, despite having equal levels of the underlying characteristic being measured. For example, younger and older patients with equal levels of disability may respond systematically differently to a self-care item such as getting dressed. DIF can be detected both statistically and graphically. In the current analysis, DIF was tested by age, gender, years of education and disease duration.

Reliability

An estimate of the internal consistency reliability of the ICF item sets was tested by both Cronbach's alpha[19] and person separation index (PSI) from the Rasch analysis.[20] The PSI is equivalent to Cronbach's alpha. Usually a reliability of 0.70 is required for analysis at the group level, and values of 0.85 and higher for individual use.[21]

External construct validity

The external construct validity was assessed by testing for expected associations of ICF item sets with RMDQ and SF-36 through the process of convergent construct validity.[22] The degree of associations with these outcome measures was analyzed by Spearman's correlation coefficient.

Sample size and statistical software

For the Rasch analysis, a sample size of 100 patients will estimate item difficulty with alpha of 0.05 to within ±0.39 logits.[23] This sample size is also sufficient to test for DIF where at alpha of 0.05 a difference of 0.39 within the residuals can be detected for any two groups with beta of 0.20. Bonferroni correction was applied to both fit and DIF statistics due to the multiple testing.[24] Statistical analysis was undertaken with SPSS for Windows version 11.5, (SPSS Inc., Chicago, Illinois, USA), Rasch analysis with RUMM2020 package.[25]

RESULTS

Patient characteristics

The mean disease duration was 88.4 months (median: 24, minimum-maximum: 1-600 months). Twenty nine percent of the patients were employed and the rest were retired (30%), housewives (35%) or unemployed (6%). The scores of patients on RMDQ and SF-36 are shown in table 1.

Internal construct validity

"Body functions and body structures" component

The Rasch analysis of this component was performed after six extreme items (b180, b260, b715, b720, b735, b750) of BF were removed. Thus, starting with the remaining 18 items, six BF and two BS items displayed disordered thresholds, necessitating collapsing of response categories. Following this, "b730-Muscle power functions", "s740-Structure of pelvic region" and "s750-Structure of lower extremity" were removed due to lack of fit, DIF by age and DIF by gender respectively. After this, the remaining 15-item BF and BS set were found to fit the model (given a Bonferroni adjustment fit level of 0.003; Table 2). Overall mean item fit residual

Table 1. Scores of patients on outcome measures (n=100)			
Scales/subscales	Mean±SD	Median	Min.-Max.
RMDQ-Total	12.9±5.6	13	0-23
SF-36 Physical functioning	48.3±18.9	50	10-89
SF-36 Role-physical	34.0±42.4	0	0-100
SF-36 Bodily pain	39.7±19.4	41.3	0-90
SF-36 General health	59.8±26.0	62.5	0-100
SF-36 Vitality	53.4±20.2	52.0	0-100
SF-36 Social functioning	56.3±38.5	62.5	0-100
SF-36 Role-emotional	54.9±30.6	55	0-100
SF-36 Mental health	60.3±20.1	62	0-100
SF-36 Physical health component	37.8±8.9	35.8	17.1-56.8
SF-36 Mental health component	44.6±10.5	46.1	20.6-69.1

SD: Standard deviation; Min.: Minimum; Max.: Maximum; RMDQ: Roland-Morris disability questionnaire; SF-36: Short Form-36.

Table 2. Fit of the body functions and body structure item set to partial credit model (15 items)

ICF code	ICF category title	Location	SE	Individual item fit residual	p^*
b126	Temperament and personality functions	−0.344	0.131	−0.981	0.672
b130	Energy and drive functions	−1.551	0.122	−1.211	0.078
b134	Sleep functions	−0.922	0.163	0.368	0.379
b152	Emotional functions	−1.220	0.121	0.095	0.822
b280	Sensation of pain	−1.285	0.159	−0.181	0.728
b455	Exercise tolerance function	−1.086	0.140	−0.572	0.600
b620	Urination functions	0.910	0.273	0.870	0.033
b640	Sexual functions	0.696	0.260	−0.222	0.741
b710	Mobility of joint functions	3.020	0.922	0.247	0.169
b740	Muscle endurance functions	2.898	0.888	0.205	0.189
b770	Gait pattern functions	−1.974	0.115	0.977	0.512
b780	Sensations related to muscles and movement functions	−1.575	0.120	0.688	0.471
s120	Spinal cord and related structures	−1.475	0.157	−0.246	0.382
s760	Structure of trunk	1.436	0.244	0.148	0.839
s770	Additional musculoskeletal structures related to movement	2.474	0.569	−0.106	0.771

ICF: The International Classification of Functioning, Disability and Health; PCM: Partial credit model; ICF: Disability and health; SE: Standard error; *: p value obtained from individual item fit statistic (chi-square).

was 0.005 (SD 0.619) and mean person fit residual was −0.165 (SD 0.561). Item trait interaction was non-significant, supporting the invariance of items [chi-square=29.87 (df=30), p=0.472]. The PSI (reliability) was good (0.77) indicating the ability of this item set to differentiate more than three groups of patients.[20] Although mean person location of −2.573 was less than that of item (0), the targeting of the items to the patients was good (Figure 1). All items were free of DIF by age, gender, years of education and disease duration.

Finally, using the PCA of residuals obtained from PCM, taking the highest positively and negatively correlated items to the first residual PC to make two subsets, no significant difference in person estimates (t=8.0%; 95% CI 3.7%-12.3%) was found between the two subsets supporting the unidimensionality of the 15-item BF and BS item set. When the assumption of local independence was examined, there was no pair of items which had a residual correlation of 0.30 or more.

"Activities and Participation" component

After removing two extreme items (d455, d859), Rasch analysis was performed on 27 items. Nineteen items displayed disordered thresholds, necessitating the collapsing of response categories. Following this, six
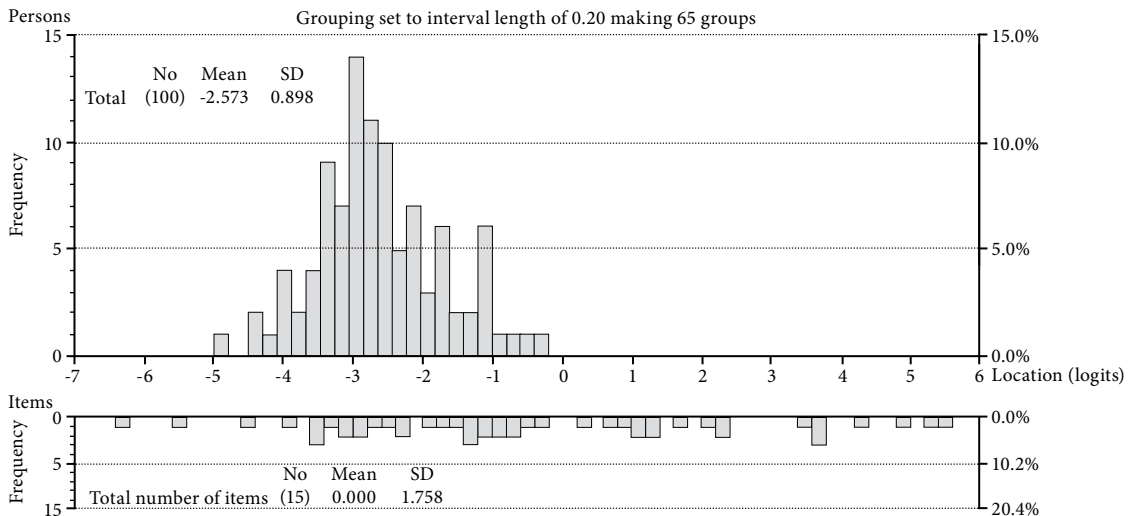

Figure 1. Person-item threshold map of BF and BS item set.

Table 3. Fit of the activity participation item set to partial credit model (21 items)

ICF code	ICF category title	Location	SE	Individual item fit residual	p*
d240	Handling stress and other psychological demands	−0.408	0.168	−0.775	0.117
d410	Changing basic body position	−1.770	0.123	−0.554	0.186
d415	Maintaining a body position	−2.016	0.135	−0.384	0.946
d420	Transferring oneself	−0.005	0.130	1.484	0.482
d430	Lifting and carrying objects	−1.901	0.154	1.303	0.366
d445	Hand and arm use	1.240	0.278	0.460	0.320
d450	Walking	−1.678	0.122	0.277	0.750
d460	Moving around in different locations	1.312	0.210	−0.791	0.157
d470	Using transportation	−0.503	0.168	0.445	0.333
d475	Driving	0.105	0.234	−0.066	0.620
d510	Washing oneself	1.388	0.219	−0.757	0.711
d530	Toileting	0.985	0.197	0.797	0.805
d540	Dressing	1.047	0.192	0.190	0.931
d570	Looking after one's health	1.445	0.242	−1.171	0.215
d620	Acquisition of goods and services	−0.776	0.169	−0.177	0.035
d640	Doing housework	−1.755	0.117	−0.125	0.721
d650	Caring for household objects	−0.914	0.182	0.410	0.878
d660	Assisting others	−0.071	0.168	−0.702	0.648
d760	Family relationship	1.299	0.239	−0.405	0.669
d850	Remunerative employment	2.139	0.523	0.915	0.036
d910	Community life	0.837	0.177	−0.506	0.502

ICF: The International Classification of Functioning, Disability and Health; SE: Standard error; *: p value obtained from individual item fit statistic (chi-square).

items, "d770-Intimate relationship", "d920-Recreation and leisure", "d710-Basic interpersonal interactions", "d630-Preparing meals", "d845-Acquiring, keeping and terminating a job" and "d465-Moving around using equipment" were removed due to response-dependency by also considering the clinical relevance of the items with the condition. Fit to the model for the remaining 21-item AP set was satisfactory (given a Bonferroni adjustment fit level of 0.002; Table 3). Overall mean item fit residual was −0.006 (SD 0.730) and mean person fit residual was −0.084 (SD 0.806). Item trait interaction was non-significant supporting the invariance of items [chi-square=42.59 (df=42), p=0.445]. The PSI was good (0.87) indicating the ability of this item set to differentiate more than four groups of patients.[20] Although mean person location of −2.231 was less than
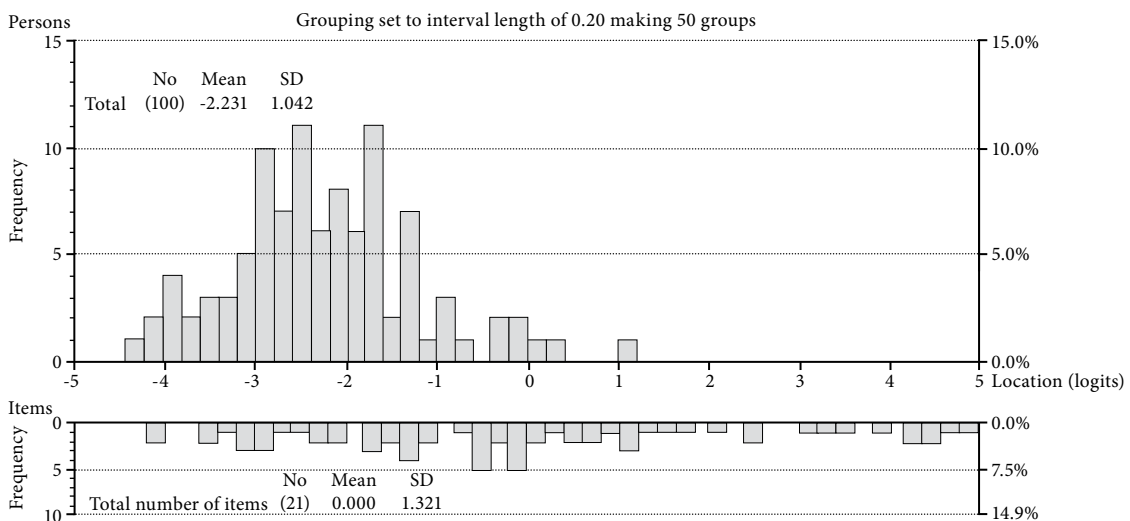

Figure 2. Person-item threshold map of activity participation item set.

Table 4. Correlations of body functions and body structure and activity participation item sets with RMDQ and SF-36®

Scales / subscales	Rasch BF and BS scores	Rasch AP scores
RMDQ-total	0.421***	0.588***
SF-36 Physical functioning	−0.472***	−0.541***
SF-36 Role-physical	−0.256*	−0.371***
SF-36 Bodily pain	−0.356***	−0.471***
SF-36 General health	−0.276**	−0.416***
SF-36 Vitality	−0.362***	−0.455***
SF-36 Social functioning	−0.189	−0.218*
SF-36 Role-emotional	−0.381***	−0.234*
SF-36 Mental health	−0.327**	−0.350***
SF-36 Physical health component	−0.357***	−0.524***
SF-36 Mental health component	−0.303**	−0.252*

RMDQ: Roland-Morris disability questionnaire; SF-36: Short Form-36; BF: Body functions; BS: Body structure; AP: Activity participation; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$.

that of item (0), the targeting of the items to the patients was good (Figure 2). All items were free of DIF by age, gender, years of education and disease duration.

Finally, using the PCA of residuals obtained from PCM, taking the highest positively and negatively correlated items to the first residual PC to make two subsets, no significant difference in person estimates (t=8.0%; 95% CI 3.7%–12.3%) was found between the two subsets supporting the unidimensionality of this item set. When the assumption of local independence was examined, there was no pair of items which had a residual correlation of 0.30 or more.

"Environmental factors" component

Rasch analysis was performed after 11 extreme items (e120, e255, e360, e455, e460, e465, e550, e570, e575, e585, e590) were removed. As 11 of the remaining 14 items displayed disordered thresholds, the relevant categories were collapsed for these items. Following this, "e410-Individual attitudes of immediate family members" and "e425-Individual attitudes of acquaintances, peers, colleagues, neighbours and community members" were removed due to lack of fit. Then the remaining 12 items were found to fit the model (given a Bonferroni adjustment fit level of 0.004) with an overall mean item fit residual of −0.117 (SD 0.493) and mean person fit residual of −0.370 (SD 0.873). Item trait interaction was non-significant supporting the invariance of items (chi-square=45.44 (df=24), p=0.005). The PCA of residuals obtained from PCM supported the unidimensionality of this item set (t=9.0%; 95% CI 4.7%-13.3%). When the assumption of local independence was examined, there was no pair of items which had a residual correlation of 0.30 or more. All items were free of DIF by age, gender,

years of education and disease duration. However, the PSI (reliability) was very low (0.29) indicating the inability of this item set to differentiate two groups of patients.[20] As the EF item set with the remaining 12 out of 25 items did not meet the assumptions of the Rasch analysis in terms of reliability (very low PSI), this item set was omitted in further analysis.

Reliability

Reliabilities of both the '"BF and BS" and AP item sets' were good, with Cronbach's alpha of 0.77 and 0.91, and PSI of 0.77 and 0.87 respectively.

External construct validity

Correlations of BF and BS and AP item sets with the RMDQ and the SF-36 are presented in table 4. The highest correlations were found with SF-36 physical functioning and RMDQ. As expected, correlations of AP item set with RMDQ and physical subsections of SF-36 were at a moderate level and higher than that of the BF and BS item set.

DISCUSSION

Standardized assessments are used widely in health care, both in clinical and research contexts. The Rasch model[10] is the current standard for the evaluation and development of assessment tools or scales delivering metric quality outcomes.[14,26] It provides not only a transformation of an ordinal score into a linear, interval-level variable but also confirms the internal construct validity of such assessments. A key characteristic of an ordinal level assessment tool is that the distances between the raw score points are unequal and mathematical calculations are invalid.[27] In contrast, an interval scale has equal interval units

which can support mathematical operations such as the calculation of summed, or change score. Thus, Rasch analysis allows for a unified approach to the construction and the internal construct validity of such assessments through the testing of its assumptions and additional aspects such as ordering of response categories and invariance of items across groups (DIF).[27]

The present study has investigated the psychometric properties of components of the ICF Comprehensive Core Set for LBP. The results of Rasch analysis indicated that it was possible to create unidimensional and robust item sets for the assessment of BF and BS and AP. However, the EF item set was not found to have sufficient reliability to be used as a measurement tool.

There was an earlier report which also explored the ICF core set for LBP by Rasch analysis in terms of construct dimensionality.[28] That paper which included 118 patients analyzed BF and BS items separately and found that the BS item set did not meet the assumptions of Rasch analysis. It showed that the remaining items of the BF set fit to the model after combining response categories. In the present study, we analyzed the BF and BS items together and, after modification of response categories and exclusion of some items, we were able to make up a unidimensional, robust 15-item BF and BS set including three items from BS and 12 items from BF. Six extreme BF items showing floor effect had to be excluded. Although the number of BF items included in our final item set was low compared with the other study, this item set was rational and more comprehensive allowing for the assessment of both BF and BS.

We were also able to create a unidimensional, robust 21-item AP set after the collapsing of response categories and the exclusion of eight items either for being extreme or showing misfit. This result differed from the earlier study where they had to make two different item sets as the whole AP set did not meet the requirements of the Rasch model.

Regarding the EF component, it was impossible in this study to make up a reliable and valid EF item set according to the assumptions of the Rasch model. However, Roe et al.[28] were able to find a 15-item EF set (out of 29 items) measuring a single underlying construct. The discrepant results between the two studies may be partly due to the difference of the response categories of the EF items in two analyses as we made the analysis on the original nine level category

whereas the earlier study rescored the response categories such that "barriers", "neither barrier nor facilitator" and "facilitators" were scored as 0, 1 and 2 respectively. Environmental differences between the two countries, Norway and Turkey, may also have affected results. A considerable number of items had to be excluded from the EF set as they were rated as "neither barrier nor facilitator" by most of the patients. Finally, another explanation for the discrepancy of the results for the EF and other components might be the dissimilarity of the characteristics of the samples in the two studies. Our patients were relatively less disabled as extreme items showing floor effect had to be excluded from the analyses.

Besides internal construct validity, expected associations with a physical disability and health-related quality of life confirmed external construct validity of the 15-item BF and BS set and 21-item AP set.

There are a number of limitations to the study. The first one is the sample size which only gives a certain degree of precision to item and person location, although when well targeted, this sample size will give an estimate of item difficulty to within 0.39 logits. Given the Rasch model allows an adaptation to interval scaling, a nomogram giving the exchange rate between the raw score and latent interval scale estimate would have been useful. However, this does require a larger sample size (e.g. 250 cases or 20 times the number of items, whichever is the larger) and so will have to wait until larger replications are undertaken. The collapsing of categories also impedes the production of the exchange rate as this will require further evidence and consensus of scoring options. Another limitation is the non-heterogeneity of our patient population which includes mostly non-employed females with considerably less disability levels.

In conclusion, it was possible to derive unidimensional, robust BF and BS and AP item sets from the ICF comprehensive core set for the assessment of functioning in patients with LBP. Both the 15-item BF and BS set and the 21-item AP set were found to be reliable and valid. However, these results need to be verified in larger and heterogeneous LBP patient groups and also tested for cross-cultural validity.

Declaration of conflicting interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

REFERENCES

1. Hillman M, Wright A, Rajaratnam G, Tennant A, Chamberlain MA. Prevalence of low back pain in the community: implications for service provision in Bradford, UK. J Epidemiol Community Health 1996;50:347-52.

2. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research. A proposal for standardized use. Spine (Phila Pa 1976) 1998;23:2003-13.

3. Katz NJ. Measures of adult back and neck function. Arthritis Rheum 2003;49:S43-9.

4. World Health Organization: International Classification of Functioning, Disability and Health. Geneva: ICF: 2001.

5. Cieza A, Stucki G, Weigl M, Disler P, Jäckel W, van der Linden S, et al. ICF Core Sets for low back pain. J Rehabil Med 2004;(44 Suppl):69-74.

6. Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine (Phila Pa 1976) 2000;25:3115-24.

7. Küçükdeveci AA, Tennant A, Elhan AH, Niyazoglu H. Validation of the Turkish version of the Roland-Morris Disability Questionnaire for use in low back pain. Spine (Phila Pa 1976) 2001;26:2738-43.

8. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.

9. Koçyigit H, Aydemir O, Fişek G, Ölmez N, Memiş A. Kısa form-36 (KF-36)'nın Türkçe versiyonunun güvenilirliği ve geçerliliği. Romatizmal hastalıkları olan bir grup hasta ile çalışma. İlaç ve Tedavi Dergisi 1999; 12:102-6.

10. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1960.

11. Luce RD, Tukey JW. Simultaneous conjoint measurement: a new scale type of fundamental measurement. J Math Psychol 1964;1:1-27.

12. Newby VA, Conner GR, Grant CP, Bunderson CV. The Rasch model and additive conjoint measurement. J Appl Meas 2009;10:348-54.

13. Masters GN. A Rasch model for partial credit scoring. Psychometrika 1982; 47:149-74.

14. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007;57:1358-62.

15. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 2007;46:1-18.

16. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3:205-31.

17. Wright BD. Local dependency, correlations and principal components. Rasch Meas Trans 1996;10:509-11.

18. Teresi JA, Kleinman M, Ocepek-Welikson K. Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. Stat Med 2000;19:1651-83.

19. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334.

20. Fisher WP. Reliability statistics. Rasch Measure Trans 1992;6:238.

21. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 2nd ed. New York: Oxford Medical Publications; 1995.

22. Nunnally JC. Psychometric theory. New York: McGraw-Hill; 1978.

23. Linacre JM. Sample size and item calibration stability. Rasch Meas Trans 1994;7:328

24. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995;310:170.

25. Andrich D, Lyne A, Sheridan B, Luo G. RUMM2020: Rasch Unidimensional Measurement Models Software. RUMM Laboratory Perth, Western Australia; 2003.

26. Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. Value Health 2004;7 Suppl 1:S22-6.

27. Elhan AH, Küçükdeveci AA, Tennant A. The Rasch measurement model. In: Franchignoni F, editor. Advances in Rehabilitation. Research Issues in Physical & Rehabilitation Medicine. Pavia: Maugeri Foundation Books; 2010. p. 89-102.

28. Røe C, Sveen U, Geyh S, Cieza A, Bautz-Holter E. Construct dimensionality and properties of the categories in the ICF Core Set for low back pain. J Rehabil Med 2009;41:429-37.