Rasch Analysis of the International Quality of Life Basic Data Set Version 2.0

Post, Marcel W.M.; Fellinghauer, Carolina S.; Charlifue, Susan; New, Peter W.; Forchheimer, Martin B.; Tate, Denise G.

Published in:
Archives of Physical Medicine and Rehabilitation

DOI:
10.1016/j.apmr.2022.02.018

Publication date:
2022

ACRM
AMERICAN CONGRESS OF REHABILITATION MEDICINE

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2022;103: 2120–30

ORIGINAL RESEARCH

Rasch Analysis of the International Quality of Life Basic Data Set Version 2.0

Marcel W.M. Post, PhD,[a,b] Carolina S. Fellinghauer, PhD,[c] Susan Charlifue, PhD,[d] Peter W. New, MD, PhD,[e,f,g] Martin B. Forchheimer, MPP,[h] Denise G. Tate, PhD[h]

From the [a]Center of Excellence for Rehabilitation Medicine, UMC Utrecht Brain Center, University Medical Center Utrecht and De Hoogstraat Rehabilitation, Utrecht, the Netherlands; [b]University of Groningen, University Medical Center Groningen, Department of Rehabilitation Medicine, Groningen, the Netherlands; [c]Swiss Paraplegic Research, Nottwil, Switzerland; [d]Craig Hospital, Englewood, CO; [e]Spinal Rehabilitation Service, Caulfield Hospital, Alfred Health, Melbourne, Victoria, Australia; [f]Epworth-Monash Rehabilitation Medicine Unit, Monash University, Melbourne, Victoria, Australia; [g]Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia; and [h]Department of Physical Medicine and Rehabilitation, University of Michigan, Ann Arbor, MI.

Abstract

Objective: To examine the internal construct validity of the International Spinal Cord Injury Quality of Life Basic Data Set Version 2.0 (QoL-BDS V2.0) and compare this with the internal construct validity of the original version of the QoL-BDS.

Design: International cross-sectional psychometric study.

Setting: Spinal rehabilitation units, clinics, and community.

Participants: The study involved 5 sites and 4 countries, 2 of whose primary language is not English. Each site included a consecutive sample of inpatients with spinal cord injury or disease (SCI/D) and a convenience sample of individuals with SCI/D living in the community (N=565).

Main Outcome Measures: The QoL-BDS V2.0 consists of the 3 original items on satisfaction with life as a whole, physical health, psychological health of the QoL-BDS, and an additional item on satisfaction with social life. All 4 items are answered on a 0-10 numeric rating scale. Rasch analysis was performed on versions 1.0 and 2.0 of the QoL-BDS to examine the ordering of the items' response options, item scaling, reliability, item fit, local item independence, differential item functioning, and unidimensionality.

Results: The sample included 565 participants with 57% outpatients and 43% inpatients. Mean age was 51.4 years; 71% were male; 65% had a traumatic injury, 40% had tetraplegia, and 67% were wheelchair users. Item thresholds were collapsed for ordering, and subsequent analyses showed good internal construct validity for the QoL-BDS V2.0 with a person separation reliability of 0.76 and Cronbach α of 0.81. Infit and outfit statistics ranged 0.62-0.91. No local dependencies and multidimensionality were found. Differential item functioning was observed only for country and inpatients vs outpatients but not for other participants' characteristics. Differences in internal construct validity between the 3-item and 4-item versions were minimal.

Conclusions: The results of this Rasch analysis support the internal construct validity of the QoL-BDS V2.0.

Archives of Physical Medicine and Rehabilitation 2022;103:2120–30

Spinal cord injury or disease (SCI/D) is associated with multiple impairments of body systems and reduced abilities, along with many potential secondary health conditions, leading to decreased quality of life (QOL) of the individuals involved.[1] QOL therefore is an important factor when evaluating the outcomes of rehabilitation for people with SCI/D. Many instruments to evaluate QOL among individuals with SCI/D currently are available, but most of these are relatively long and complex and were developed for use in community settings.[2] Also, with the exception of the World Health Organization Quality of Life,[3] there are few instruments that focus on this concept from a worldwide perspective.

Funded by the Craig H. Neilsen Foundation, grant application ID no. 440840.
Disclosures: none.

0003-9993/$36 - see front matter
https://doi.org/10.1016/j.apmr.2022.02.018

Therefore, there is a need for a brief QOL measure that can be used internationally and in both the initial rehabilitation of SCI/D and after people have returned to the community.

The International Spinal Cord Injury Quality of Life Basic Data Set (QoL-BDS) was developed as a brief QOL measure for use in clinical practice and research.[4,5] Its purpose is to standardize the collection and reporting of a minimal amount of information necessary to merge and compare results of published and unpublished studies on QOL. Like all basic data sets, it was designed to include a minimal number of data elements, which together can be collected in routine clinical practice, not as the measure to replace all other QOL measures.[4] Its development was based on the definition of subjective QOL as reflecting an individual's overall perception of and satisfaction with how things are in their life.[4,6] It consists of 3 items for an individual to rate their satisfaction with life as a whole, physical health, and psychological health. Previous research in inpatient and community settings provided preliminary evidence of its cross-cultural validity,[5,7] convergent and discriminant validity,[8,9] and inter-rater reliability.[10] A fourth item on satisfaction with social life was considered during the development of the QoL-BDS, to mirror the definition of health by the World Health Organization of health as a state of physical, mental, and social well-being.[11,12] Subsequently, the decision was made to exclude the item on social life to avoid overlap with the simultaneously developed International Activities and Participation Basic Data Set.[13] More recently, however, results of an international cognitive interview study made the importance of social life as a QOL domain clear, which prompted a decision to add a fourth item to the QoL-BDS V2.0.[14] The 4-item QoL-BDS Version 2.0 (V2.0) showed similar internal consistency and reproducibility to what was found for the original 3-item version.[15]

The aim of our study is to extend the psychometric evidence on the QoL-BDS[5,7-10] by examining metric properties of the QoL-BDS V2.0 using Rasch analysis. Rasch analysis is used to determine if items from a questionnaire fulfill fundamental measurement assumptions.[16] If so, the scores can be transformed to an interval-level scale and thereby be usable for parametric statistical testing and longitudinal studies on QOL in clinical and research settings.[17] We hypothesized that the QoL-BDS V2.0 is sufficiently unidimensional, with ordered response categories and without misfitting items. We also hypothesized that its structure would be equivalent to that for the original QoL-BDS.

Methods

Design

This was an international cross-sectional validation study.

List of abbreviations:

AIS	**American Spinal Injury Association Impairment Scale**
ANOVA	**analysis of variance**
DIF	**differential item functioning**
LID	**local item dependence**
QOL	**quality of life**
QoL-BDS	**International Spinal Cord Injury Quality of Life Basic Data Set**
SCI	**spinal cord injury**
SCI/D	**spinal cord injury or disease**
V2.0	**Version 2.0**

Participants

The study involved QOL data from 5 sites and 4 countries: Australia, Brazil, the Netherlands, and 2 sites in the US: Colorado and Michigan. Each of the 5 study sites aimed to recruit and enroll 48 inpatients and 64 outpatients for the study. The 2 US sites were combined to represent that country for analytical purposes.

Eligibility criteria for all sites were (1) documented diagnosis of SCI/D without complete functional recovery (American Spinal Injury Association Impairment Scale [AIS] classification A-D); (2) minimum age of 18 years at the time of participation; (3) completion of informed consent. Individuals were excluded if they were unable to read and speak the native language of the country in which they lived or were unable to complete a self-report questionnaire because of cognitive limitations. For both the inpatient and outpatient samples, the study team aimed to recruit similar numbers of participants in terms of neurologic classification (all AIS D; Paraplegia AIS A, B, or C; and Tetraplegia AIS A, B, or C), age (younger than 50 years and older than 50 years), and etiology (traumatic and nontraumatic). Within the outpatient sample they also aimed to include a similar number of participants who had been discharged from their initial inpatient hospitalization for less than and more than 5 years.

Procedures

Inpatients were recruited while they were undergoing initial rehabilitation after the onset of their SCI/D at one of the participating study hospital sites. Outpatients were recruited from individuals visiting associated outpatient clinics, former participants with spinal cord injury (SCI) in other studies who agreed to be contacted, or from the hospitals' databases or registries. All participants received oral and written information about the study and provided informed consent. The QoL-BDS V2.0 was administered in person or by telephone interview by a trained researcher. The study was approved at all sites by their respective Institutional Review Boards and Ethics Committees.

Instruments

All information used in this study came from the study questionnaire developed to include the SCI QoL-BDS V2.0 and questions related to injury and demographic characteristics of participants.

The QoL-BDS V2.0 includes 4 items on the individual's satisfaction with their life as a whole, physical health, mental health, and social life. All items use a time frame of the past 4 weeks and a 0-10 numeric rating scale, with higher scores indicating better QOL.[4] The QoL-BDS was developed in English, and this version was used in the US and Australia and was translated into Dutch and Brazilian Portuguese following the recommendations of the International SCI Data Sets project.[7,8,18]

Other variables used in the current analyses included country, age, sex, years of education, level of SCI/D (tetraplegia or paraplegia), and mobility (wheelchair user or walker).

Statistics

A polytomous version of the Rasch model, called the partial credit model, was applied to test the metric properties of QoL-BDS V2.0 and the original, 3-item, version. The following core measurement assumptions were tested.[16,19,20]

Ordering of the items' response options

The assumption of monotonicity is satisfied if the expected item scores always increase as the person parameter (eg, health state, level of QOL) increases.

Targeting of the scale

Good targeting is achieved if mean item difficulty and mean person ("ability") parameters are similar and item difficulty thresholds cover the person parameter continuum. The person item map shows the quality of the targeting. The upper portion of the person item map displays frequencies of person parameter estimates, and the lower portion displays the location of the difficulty estimates for each test item. The x-axis represents the measurement continuum in logits. For the persons, the more negative the logit value the lower the QOL and the more positive the logit value the higher the QOL. For the items, more negative logit values indicate a certain "easiness" of the items to be endorsed with better QOL. Items that are shifted toward the right of the measurement continuum, with higher positive logit values, are "more challenging" items for the sample and are less "easily" endorsed for better QOL and correspond to lower satisfaction ratings. With good targeting, the QOL levels histogram and difficulty estimates ideally should be centered on the same point and show a considerable overlap.

Model reliability

The reliability was evaluated using several statistical coefficients. The Cronbach α is a measure of the internal consistency of the data.[21] The Person Separation is used to classify individuals, while Item Separation is used to verify the item hierarchy.[17,22] Two ways to formalize the separation exist, the Separation Indices and the Separation Coefficients. For the Person Separation Index, also called Person Separation Reliability, as with Cronbach α, values above 0.75 indicate good reliability, and values above 0.9 indicate very good reliability of the scale. A low Item Separation Index means that the sample is not big enough to locate the items on the latent variable. The Person and Item Coefficients describe the "true" spread of items or individuals along the measurement continuum and reflect the number of distinct strata in which the sample or items can be divided. A person separation coefficient of 1.5 represents an acceptable level of separation.[22]

Item and person fit

Infit and outfit statistics determine the quality of the item and person fit. Infit and outfit are mean-square residual summary statistics that can range from 0 to infinity. Unlike the infit statistic, outfit adjusts for outlying values by correcting for the residual variance. Ideally, the item infit and outfit statistics are close to 1. Values <0.8 are indicative of overfit.[23] Values between 0.5 and 1.5 are generally considered as productive for measurement.[22] To determine underfit in the items, this study further applied the sample size−adjusted cutoff values for the item infit and outfit. Specifically, outfit values >1.25 and infit values >1.08 were considered indicative for item that underfit and suggested lack of discrimination.[23]

Absence of local item dependence

Local item dependence (LID) often occurs when items are redundant and measure approximately the same aspect of a construct. LID is evaluated by inspecting correlations between the Rasch residuals. Residual correlations >0.2 above the average residual correlation suggest item redundancy.[24] This approach considers LID relative to the average observed residual correlation, rather than relative to a uniform value, because the size of LID is influenced by the number of items, the sample size, and the number of response categories. While the LID analysis focuses on detecting pairwise correlated items, it is worth mentioning that clusters of residual correlations and highly negative correlations may indicate scale multidimensionality.[25]

Absence of differential item functioning

The analysis of uniform differential item functioning (DIF) flags variables that lack invariance in item difficulty across subgroups.[26,27] An analysis of variance (ANOVA) was used to test for DIF in terms of country, setting (inpatient or outpatient), age (younger or older than 50 years), sex, injury level (paraplegia or tetraplegia), mobility level (wheelchair user or walker), and etiology (traumatic or nontraumatic). Benjamini-Hochberg correction for multiple testing was applied to minimize the false discovery rate.[28] To explore the DIF results further, Rasch trees (ie, partial credit trees for polytomous items) were applied to variables flagged for DIF in the ANOVA analysis. Rasch trees are a hypothesis-generating approach that allows for testing several DIF variables at once and determines main and interacting DIF effects.[29,30]

Unidimensionality

Unidimensionality was assessed with an eigenvalue decomposition of the residual correlations, which is one way to perform a principal component analysis.[31] Second eigenvalues <1.4 supported unidimensionality.[32]

The metric analyses were performed with the software R version 4.0.2,[33,a] more specifically the R packages *eRm* for the Rasch analysis,[34] *iarm* for the DIF analysis with ANOVA,[35] and *psychotree*[30] for the DIF analysis with Rasch trees.

Results

Descriptive statistics

The complete sample contained 565 participants with SCI/D, of whom 322 (57%) were outpatients. The descriptive statistics for the entire sample and stratified by country and setting are shown in table 1 and the scale item descriptive statistics in appendix 1. In the entire sample, the mean age was 51.4 years, 71.2% of participants were male, 65% had SCI, 40.4% had tetraplegia, and 66.7% were wheelchair users. There were some differences in sample composition between countries. Most notably, the cohorts in Australia and the Netherlands included higher proportions of participants with nontraumatic etiology and lower proportions of wheelchair users than the United States and Brazil (see table 1.1). Age groups differed significantly across settings, with significantly fewer participants older than 50 years in Brazil (37.5%). The proportion of wheelchair users (>75%) and of traumatic injuries (>78%) was also significantly higher in the US and Brazilian samples. The outpatient setting counted significantly more traumatic injuries (73.6%). The n=113 nontraumatic etiologies in the inpatient setting included the following causes: 41.6% degenerative spine conditions, 15.0% vascular causes, 14.2% bacterial or viral infections, 8.0% malignant tumor, and 7.1 % benign tumors. In the outpatient cohort, n=85 participants with nontraumatic etiologies included 25.9% with degenerative spine conditions, 17.7%

Table 1 Sociodemographic descriptive statistics for the study population as a whole and stratified by country and by setting

		Country					Setting		
Variable	Overall	Australia	Brazil	Netherlands	US	Sign.	Inpatient	Outpatient	Sign.
Population, n (%)	565	102 (18.1)	112 (19.8)	114 (20.2)	237 (41.9)		243 (43)	322 (57)	
Site, n (%)									
Australia	102 (18.1)	x	x	x	x		38 (15.6)	64 (19.9)	
Brazil	112 (19.8)	x	x	x	x		48 (19.8)	64 (19.9)	
Netherlands	114 (20.2)	x	x	x	x		50 (20.6)	64 (19.9)	
US	237 (41.9)	x	x	x	x		107 (44.0)	130 (40.4)	
Setting=outpatient, n (%)	322 (57.0)	64 (62.7)	64 (57.1)	64 (56.1)	130 (54.9)		x	x	
Age group >50 y, n (%)	309 (54.7)	71 (69.6)	42 (37.5)	79 (69.3)	117 (49.4)	*	136 (56.0)	173 (53.7)	
Age (y), mean ± SD	51.4 (16.4)	57.6 (14.8)	46.5 (15)	58 (14.9)	47.9 (16.6)	* †	51.2 (17.4)	51.6 (15.6)	
Sex, n (%)									
Male	402 (71.2)	56 (54.9)	89 (79.5)	77 (67.5)	180 (75.9)		175 (72.0)	227 (70.5)	
Female	162 (28.7)	45 (44.1)	23 (20.5)	37 (32.5)	57 (24.1)		67 (27.6)	95 (29.5)	
Other	1 (0.2)	1 (1.0)	0 (0.0)	0 (0.0)	0 (0.0)		1 (0.4)	0 (0.0)	
Mobility=wheelchair user, n (%)	377 (66.7)	48 (47.1)	85 (75.9)	59 (51.8)	185 (78.1)	*	177 (72.8)	200 (62.1)	†
Lesion type, n (%)						*			
Paraplegia	336 (59.5)	72 (70.6)	85 (75.9)	73 (64.0)	106 (44.7)		136 (56.0)	200 (62.1)	
Tetraplegia	228 (40.4)	30 (29.4)	27 (24.1)	41 (36.0)	130 (54.9)		106 (43.6)	122 (37.9)	
Missing	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.4)		1 (0.4)	0 (0.0)	
Type=traumatic, n (%)	367 (65.0)	39 (38.2)	88 (78.6)	53 (46.5)	187 (78.9)	*	130 (53.5)	237 (73.6)	*
Years of education, n (%)						*			
0-9	74 (13.1)	9 (8.8)	58 (51.8)	6 (5.3)	1 (0.4)		39 (16.0)	35 (10.9)	†
10-16	365 (64.6)	74 (72.5)	47 (42.0)	61 (53.5)	183 (77.2)		163 (67.1)	202 (62.7)	
>16	123 (21.8)	19 (18.6)	4 (3.6)	47 (41.2)	53 (22.4)		38 (15.6)	85 (26.4)	
Missing	3 (0.5)	0 (0.0)	3 (2.7)	0 (0.0)	0 (0.0)		3 (1.2)	0 (0.0)	

Abbreviation: Sign., significance.
* $P<.001$.
† $P<.05$.

with vascular causes, 17.7% resulting from bacterial or viral infections, 9.4% from benign tumors, and 7.1% from spina bifida. Scores on the QOL items also differed significantly across countries and setting (see appendix 1)

Rasch analysis

Results of the Rasch analyses before and after collapsing response options because of disordered response thresholds are displayed in the top half and the bottom half of table 2 and in figs 1 and 2, respectively. The lower part of table 2 shows the fit statistics of the 3-item version, analyzed separately, to see the changes with the additional item.

Ordering of response options

All items showed disordered thresholds, for example 1.2, 5.4 for the item social life, whereas an ascending order is expected (see fig 1). Monotonicity could be achieved through collapsing the 11-response scale into 4 or 5 response levels. Figure 1 shows the disordering of the response options across all 4 QOL items, whereas such disordering is not present in fig 2.

Targeting of the scale

The sample presents participants with higher levels of QOL and a few with lower QOL than what the scale can capture. In general, the scope of the measurement improved after collapsing of response options. Before collapsing of response options, the items were covering 48% of the person parameter range. After collapsing, almost 60% can be assessed. The thresholds cover a range of 2.63 logits at start and 4.71 logits after collapsing, for a person parameter range of 5.49 and 8.08 logits, respectively.

Model reliability was good

Both Cronbach α and the person separation index were above 0.8, indicating good reliability of the scale. The item coefficients and person coefficients both showed good levels of separation. About 5 strata, that is, QOL levels, can be determined in the study sample before the collapsing of response options and about 3 are determined afterward.

Item and person fit

The items presented no outfit values above 1.25 and no infit values above 1.08. On the other hand, low outfit and infit values indicate item overfit (see table 2). Items with infit and outfit statistics <0.8 are overdiscriminating, and responses to these items are more predictable than what the model expects. The items are still "productive for measurement" and not degrading the scale. The person fit was good, with no infit and outfit values above 1.5 and no values below 0.5.

Local item dependence

The analysis-specific cutoff points for LID are given in table 3, that is, at the start and final, after collapsing of reversed thresholds, for the QoL-BDS and QoL-BDS V2.0. The results support the absence of LID because no single residual correlation was above the analysis-specific cutoffs.

Table 2 Results of the Rasch analysis of the QoL-BDS 4-item and 3-item version including item fit and difficulty, ordering of thresholds, local item dependency, differential item functioning, targeting, and reliability at start of the analysis and after adjustments

QoL-BDS V2.0 4-item Version	Start	Item/Testlet	Outfit	Infit	Item Difficulty	Response Coding	Disordered Thresholds	LID	DIF
		Life as a Whole	0.62	0.63	0.38	0-1-2-3-4-5-6-7-8-9-10	Yes	No	
		Physical Health	0.87	0.88	0.44	0-1-2-3-4-5-6-7-8-9-10	Yes	No	
		Psychological Health	0.69	0.70	0.16	0-1-2-3-4-5-6-7-8-9-10	Yes	No	Site, situation
		Social Life	0.88	0.89	0.28	0-1-2-3-4-5-6-7-8-9-10	Yes	No	Site
			Mean	SD	Min; Max	Mean Residuals	SD Residuals		
	Targeting	Difficulty	0.32	0.74	−0.66; 1.97	0.25	0.09		
		Ability	0.75	0.75	−1.88; 3.61	0.33	0.13		
			Index	Coefficient	Strata				
	Reliability	Difficulty	0.87	2.81	3.09				
		Ability	0.78	2.07	4.08				
			α						
		Test	0.82						
		Item/Testlet	Outfit	Infit	Item Difficulty	Response Coding	Disordered Thresholds	LID	DIF
	Final	Life as a Whole	0.62	0.64	0.57	0-0-0-1-1-2-2-3-3-4-4	No	No	
		Physical Health	0.85	0.86	0.88	0-0-1-1-2-2-3-3-4-4-5	No	No	
		Psychological Health	0.70	0.71	0.19	0-0-0-1-1-2-2-3-3-4-4	No	No	Site, situation
		Social Life	0.90	0.91	0.59	0-0-1-1-2-2-3-3-4-4-5	No	No	Site
			Mean	SD	Min; Max	Mean Residuals	SD Residuals		
	Targeting	Difficulty	0.58	1.40	−1.2; 3.51	0.18	0.04		
		Ability	1.14	1.31	−3.05; 5.03	0.62	0.16		
			Index	Coefficient	Strata				
	Reliability	Difficulty	0.98	7.87	10.82				
		Ability	0.76	1.86	2.82				
			α						
		Test	0.81						
QoL-BDS 3-item version	Start	Item/Testlet	Outfit	Infit	Item Difficulty	Response Coding	Disordered Thresholds	LID	DIF
		Life as a Whole	0.61	0.61	0.44	0-1-2-3-4-5-6-7-8-9-10	Yes	No	
		Physical Health	0.75	0.76	0.50	0-1-2-3-4-5-6-7-8-9-10	Yes	No	
		Psychological Health	0.68	0.69	0.20	0-1-2-3-4-5-6-7-8-9-10	Yes	No	Site, situation
			Mean	SD	Min; Max	Mean Residuals	SD Residuals		
	Targeting	Difficulty	0.38	0.87	−0.71; 2.19	0.26	0.10		
		Ability	0.82	0.84	−1.75; 3.67	0.40	0.13		
			Index	Coefficient	Strata				
	Reliability	Difficulty	0.90	3.24	4.65				
		Ability	0.75	1.86	2.82				
			α						
		Test	0.80						
		Item/Testlet	Outfit	Infit	Item Difficulty	Response Coding	Disordered Thresholds	LID	DIF
	Final	Life as a Whole	0.65	0.65	0.68	0-0-0-1-1-2-2-3-3-4-4	No	No	
		Physical Health	0.73	0.74	1.05	0-0-1-1-2-2-3-3-4-4-5	No	No	
		Psychological Health	0.71	0.71	0.25	0-0-0-1-1-2-2-3-3-4-4	No	No	Site, situation
			Mean	SD	Min; Max	Mean Residuals	SD Residuals		
	Targeting	Difficulty	0.69	1.71	−1.29; 4.09	0.19	0.04		
		Ability	1.26	1.53	−2.93; 5.30	0.78	0.17		
			Index	Coefficient	Strata				
	Reliability	Difficulty	0.99	8.99	12.31				
		Ability	0.73	1.69	2.59				
			α						
		Test	0.80						

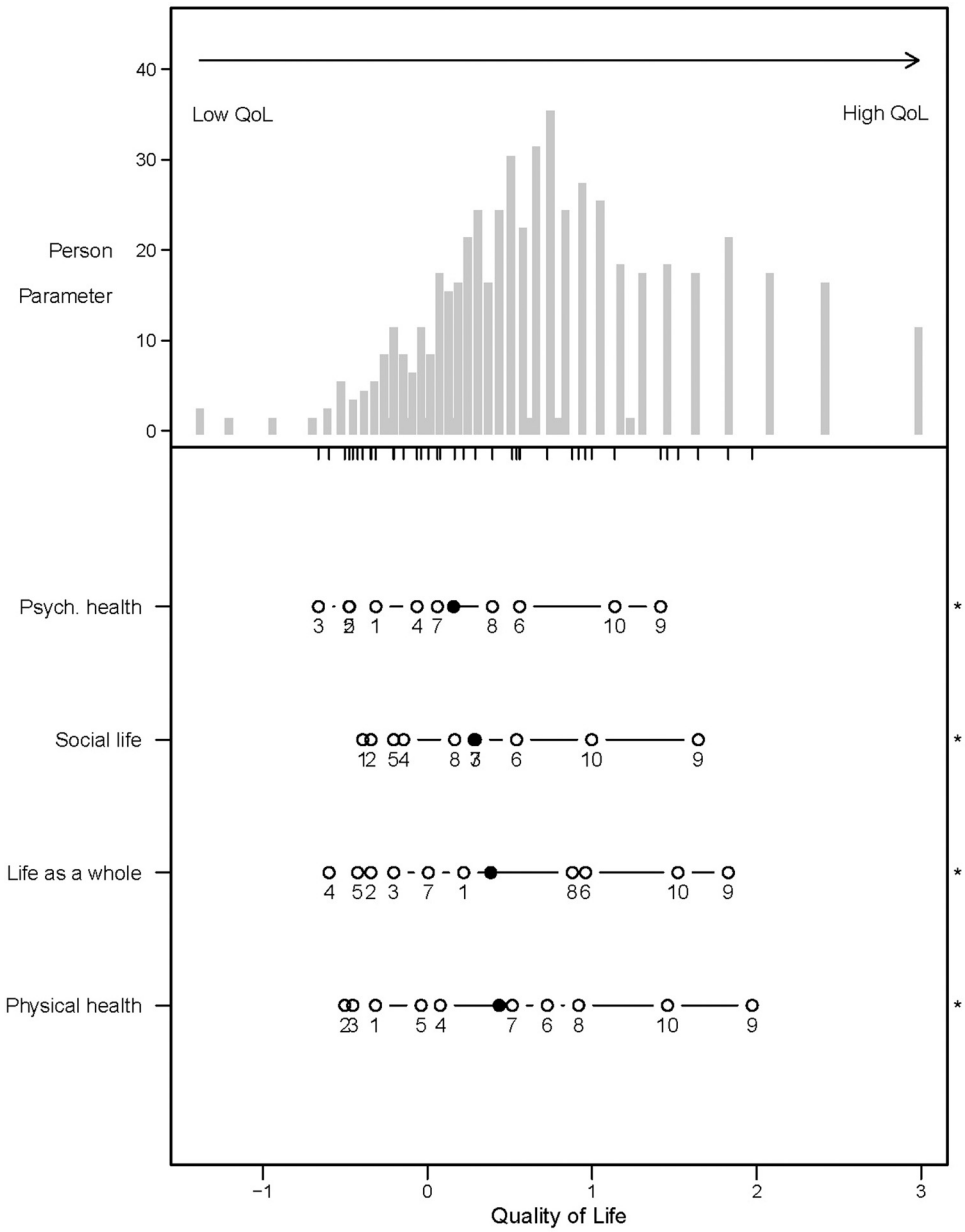

Fig 1 Person-item map and ordering of item thresholds before collapsing of response options.

DIF testing

DIF was tested for country, setting, age, sex, injury level, mobility level, and etiology. The results are shown in table 2. Participant characteristics (age, sex, lesion level, mobility level, injury etiology) were not associated with significant DIF. However, the variable "country" produced significant DIF in the items concerning Psychological Health and Social Life. DIF also was found for Psychological Health between inpatient and outpatient settings (see table 2). Further exploration with Rasch trees showed that DIF was associated, first, with the setting and, for the outpatient setting, with the country: Netherlands and Australia on the one hand and US and Brazil on the other (fig 3). The 3 response profile plots (or Rasch tree nodes) at the bottom of the fig indicate the position of the level of difficulty estimates for each of the 4 QoL-BDS V2.0 items. The item Physical Health consistently showed the highest item difficulty estimate, corresponding to lower satisfaction ratings. The difficulty estimate for the item Psychological Health was lowest, that is, satisfaction was highest, compared with the other items in all settings, especially for inpatients in the Netherlands and Australia. Relative to the other items, the item Life as a Whole received significantly higher ratings, and the item Social Life received lower ratings among inpatients. For outpatients, Social Life received higher ratings than Life as a Whole. Appendix 2 provides the infit and outfit statistics, as well as the location and difficulty thresholds in the terminal Rasch tree nodes. The item fit statistics in each Rasch tree's final node did not show any misfitting item.

Unidimensionality

Finally, the dimensionality analyses provided second eigenvalues <1.4 (ie, 1.30), supporting that the QoL-BDS V2.0 measures a single latent construct (see table 3).

Analysis of the metric properties of the original 3-item version of the QoL-BDS provided very similar results regarding item

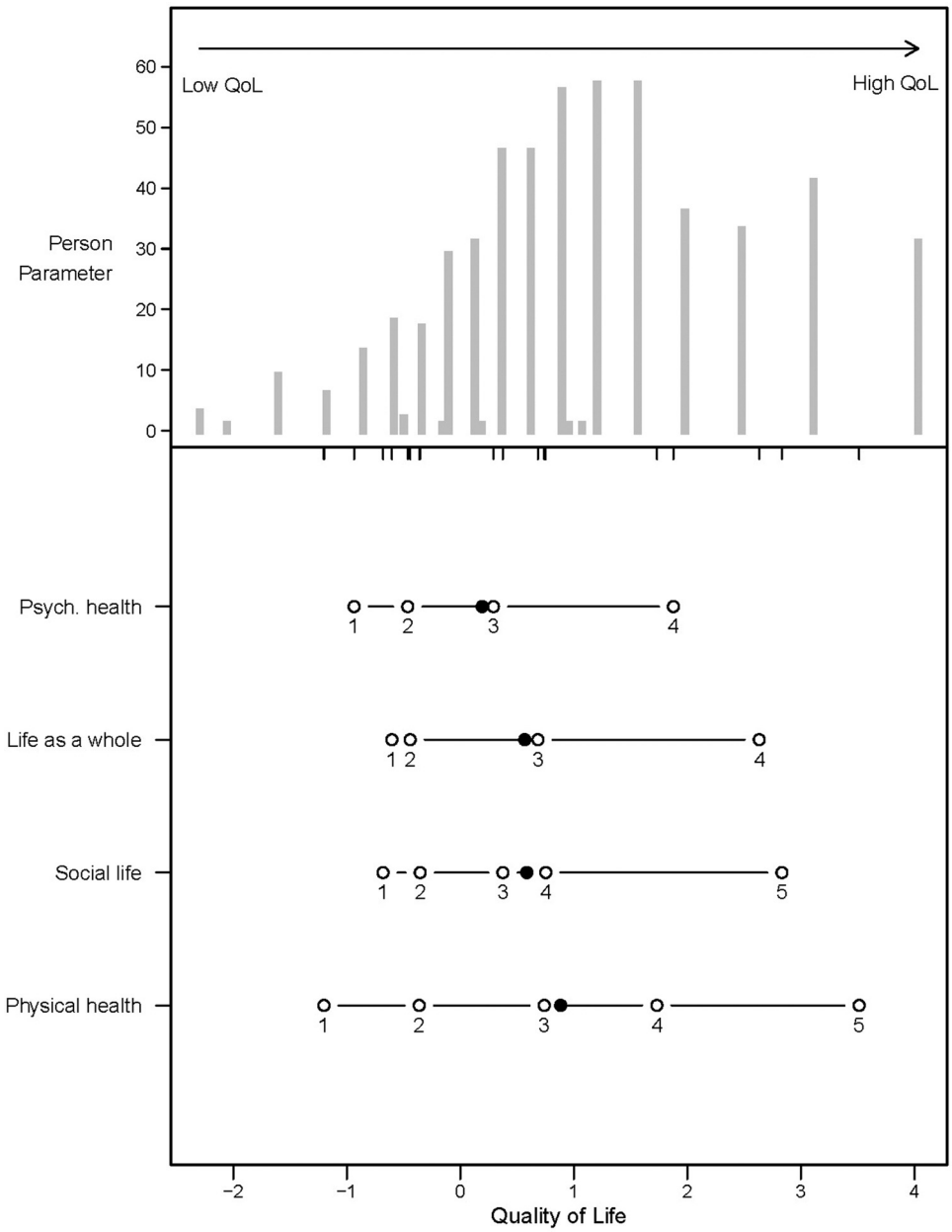

Fig 2 Person-item map and ordering of item thresholds after collapsing of response options.

disordering, DIF, and item fit, with a similar tendency toward item overfit (see table 2). The reliability of the 4-item version was somewhat higher (4-item QoL-BDS V2.0: Person Separation Index=0.82, α=0.81; 3-item QoL-BDS: Person Separation Index=0.80, α=0.77). The association between the level of QOL issued by the BDS-QoL and the BDS-QoL V2.0 after collapsing of the response options was moderate. The correlation of r=0.64 indicated a shared variance of only 41%. Given the unidimensionality of the 3- and 4-item versions, this may indicate that the additional item not only adds a supplementary but also a complementary perspective to the QOL measure. Table 4 provides disaggregated descriptive statistics of the person parameter estimates, that is, QOL levels, for the 3- and 4-item version of the QoL-BDS. The P values are corrected for the false discovery rate. The 4-item version evens up differences in QOL life observed in the shorter version of the instrument. On the other hand, differences in the perceived QOL by education level become apparent with the additional Social Life item.

Discussion

This large international study provides the first analysis of the internal construct validity of the QoL-BDS V1.0 and V2.0 using a contemporary psychometric methodology. Both our hypotheses were supported. In general, both the original QoL-BDS and the QoL-BDS V2.0 showed good internal construct validity as a 1-dimensional scale. The QoL-BDS V2.0 covers the main domains of QOL (overall, physical, mental, social), and the results indicate that each item is relevant in contributing information, as demonstrated by the absence of LID, and that these items are sufficiently coherent to make up a scale with good internal construct validity.

Table 3 Residual correlations, respective cutoffs for independence, and eigenvalues for the dimensionality

QoL-BDS: Residual Correlations				
Start Cutoff: −0.30	Life as a Whole	Physical Health	Eigenvalues	
Life as a Whole	1.00		First	1.63
Physical Health	−0.53	1.00	Second	1.37
Psychological Health	−0.37	−0.59	Third	0.00
Final Cutoff: −0.29	Life as a Whole	Physical Health	Eigenvalues	
Life as a Whole	1.00		First	1.67
Physical Health	−0.56	1.00	Second	1.32
Psychological Health	−0.32	−0.60	Third	0.01

QoL-BDS V2.0: Residual Correlations					
Start Cutoff: −0.13	Life as a Whole	Physical Health	Psychological Health	Eigenvalues	
Life as a Whole	1.00			First	1.51
Physical Health	−0.26	1.00		Second	1.28
Psychological Health	−0.23	−0.35	1.00	Third	1.21
Social Life	−0.36	−0.48	−0.29	Fourth	0.00
Final Cutoff: −0.12	Life as a Whole	Physical Health	Psychological Health	Eigenvalues	
Life as a Whole	1.00			First	1.54
Physical Health	−0.25	1.00		Second	1.30
Psychological Health	−0.17	−0.32	1.00	Third	1.16
Social Life	−0.35	−0.52	−0.32	Fourth	0.00

The only findings that did not comply with the Rasch model's assumptions were (1) disordered response options that could be resolved by collapsing response options and (2) DIF for site and setting. Differences between the 3- and 4-item versions were small, and the correlation of 0.64 between the person parameter estimates of both versions suggest the fourth item enriches the scale.

The results of this Rasch analysis support the internal construct validity of the QoL-BDS V2.0 and extend preliminary findings from previous studies in the US and Brazil.[5,7] These results further concur with results of a cognitive interview study that is part of the same project but performed in different samples from all 5 sites.[14] The interviews showed that individuals with SCI/D from these 4 countries have a largely similar understanding of the concepts measured with the QoL-BDS V2.0.

That we found disordered response thresholds is not surprising given the large number of response options per item, 11. Our analysis shows that collapsing these into 5-6 response options as displayed in the bottom half of fig 2 solved this problem. Moreover, the reduction of response options also improved the targeting. This suggests that it could be possible to reduce the number of response options of the QoL-BDS V2.0, for example into a 0-4 or 0-5 scale in a future revision. We recommend further research in the future to optimize the structure of the QoL-BDS V2.0.

Median scores on the QoL-BDS V2.0 were above the midpoint of the scale (7-7.5). Similarly, the person-item maps also showed higher QOL among the participants than what the scale is targeting on average. The percentages of maximum scores, however, indicated no notable ceiling effects. The lower item difficulty score on psychological health compared with the higher item difficulty score for physical health means, unsurprisingly, that given a certain level of QOL, participants rated their psychological health more positively than their physical health.

The presence of DIF indicated that the difficulty of some items of the QoL-BDS V2.0 varied based on setting and country. The Rasch trees showed that DIF exists between the different settings and across countries. DIF does not necessarily indicate bias in measurement if a real difference could be present, and we feel the DIF we found reflects real differences in the experience of QOL, in this case between the inpatient and the outpatient setting.[36] In this case, DIF for settings could result from psychological adjustment to living with SCI, which is associated with higher scores on the Satisfaction With Life as a Whole item without a corresponding increase in scores on the Satisfaction With Physical Health item over time.[37,38] Further, DIF for countries was limited, mainly focusing on the Satisfaction with Psychological Health item. Standardized for overall level of QOL, scores on this item were somewhat higher in Australia and the Netherlands than in the US and Brazil. Differences in sample composition could have contributed to this finding because the samples in Australia and the Netherlands included higher proportions of participants with nontraumatic etiologies and lower proportions of wheelchair users than the US and Brazil. Although these characteristics did not show significant DIF, the combination of these characteristics could have contributed to this variance. Considering the similarities in the response difficulty patterns as displayed in fig 3, adjustments of the QoL-BDS V2.0 scores for DIF by settings and sites were deemed unnecessary.

Study limitations

Although data were collected in 4 countries in different parts of the world, there was no representation of low-income or lower-middle-income countries and countries from Africa, the Middle

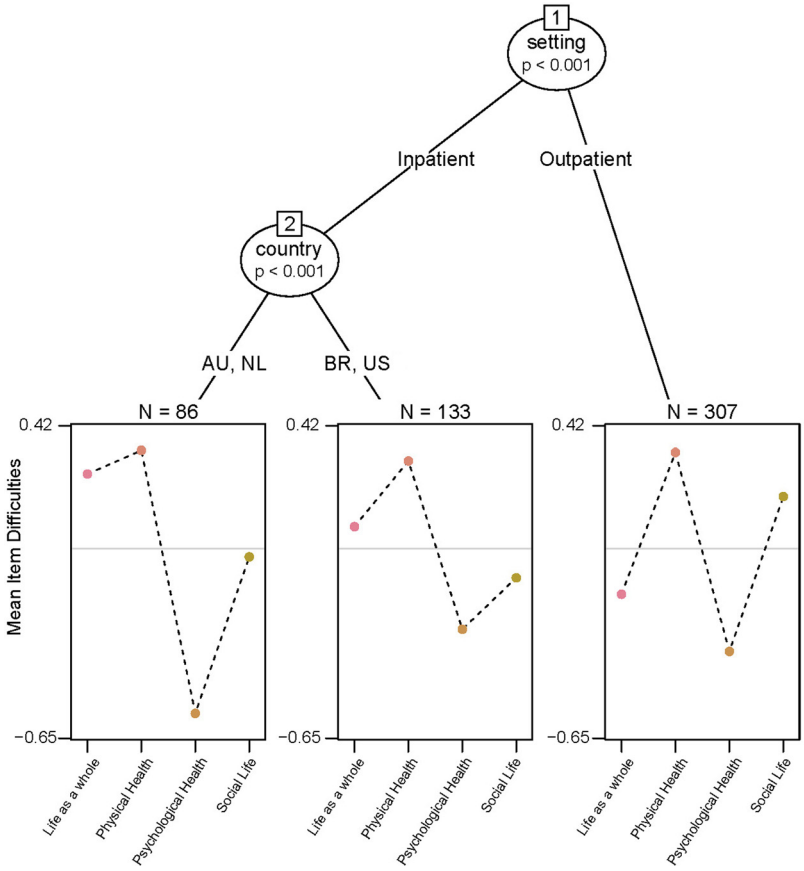

Fig 3 DIF trees, showing that item variance is related to, first, setting and, second, within inpatients to country. A lower mean item difficulty indicates an "easier" item, meaning that persons are more likely to report high QOL on this item than the other items. A higher item difficulty indicates a more challenging item, meaning that persons are less likely to score high on this item than the other items.Abbreviations: AU, Australia; BR, Brazil; NL, the Netherlands.

Table 4 Levels of QoL in the individuals, in logits, for the 3-item and 4-item version of the QoL-BDS.

		QoL-BDS: 3 Items		QoL-BDS V2.0: 4 Items	
Variable	Level	Mean ± SD	*P* Value	Mean ± SD	*P* Value
Site	Australia	0.79±1.4	.033*	1.06±1.61	.288
	Brazil	1.02±1.3		1.07±1.61	
	Netherlands	1.21±0.99		1.3±1.25	
	US	1.28±1.41		1.41±1.57	
Situation	Inpatient	0.92±1.32	.007*	1±1.56	.003
	Outpatient	1.28±1.29		1.45±1.47	
Age group	≤50 y	1.19±1.36	.489	1.31±1.6	.634
	>50 y	1.07±1.28		1.22±1.46	
Sex	Male	1.15±1.29	.548	1.3±1.52	.503
	Female	1.07±1.38		1.15±1.55	
Mobility	Walking	1.07±1.26	.548	1.21±1.49	.634
	Wheelchair user	1.16±1.35		1.29±1.54	
Lesion type	Paraplegia	1.13±1.34	.903	1.27±1.53	.814
	Tetraplegia	1.12±1.29		1.24±1.52	
Lesion type	Nontraumatic	0.87±1.23	.004*	0.95±1.43	.002*
	Traumatic	1.27±1.34		1.43±1.55	
Education level	<10 y	0.86±1.33	.090	0.76±1.59	.008*
	10-16 y	1.1±1.31		1.24±1.5	
	>16 y	1.35±1.3		1.56±1.49	

* p<.05

East, or Asia. Replication of the current study involving such countries is necessary to be able to make more universal claims about the cross-cultural validity of the QoL-BDS V2.0. Findings from this study focused exclusively on testing the internal construct or metric validity of the QoL-BDS V2.0 across sites and settings. Ongoing analyses of data collected in this study on the concurrent/divergent validity, responsiveness, and clinical utility of the QoL-BDS V2.0 will be published in due course.

Conclusions

This study supports the use of the Qol-BDS V2.0 as a brief measure of QOL that can be used in clinical settings and research. In the context of the continuing debate on the conceptualization and measurement of QOL, the QoL-BDS V2.0 is a significant step toward unifying our ability to record and report this important information.

Supplier

a. R: a language and environment for statistical computing; R Foundation for Statistical Computing.

Keywords

Psychometrics; Quality of life; Rehabilitation; Spinal cord injuries; Validation studies

Corresponding author

Marcel W.M. Post, PhD, Center of Excellence for Rehabilitation Medicine, De Hoogstraat Rehabilitation, Rembrandtkade 10, 3583TM Utrecht, the Netherlands. *E-mail address:* m.post@dehoogstraat.nl.

Acknowledgments

We thank the local research team members. Melbourne: Diana Ramirez Hernandez; Sao Paulo: Angelica Castilho, Julia Maria D'Andrea Greve, Alexandra Cristoffi, and Carla Witter; Colorado: Jennifer Coker, Abbey Welch, and Bria MacIntyre; Michigan: Constance Pines; Utrecht: Christel van Leeuwen and Aline van der Wind. We also thank Carolin Strobl for guidance on some analytical aspects of this study.

References

1. Post MW, van Leeuwen CM. Psychosocial issues in spinal cord injury: a review. Spinal Cord 2012;50:382–9.
2. Hill MR, Noonan VK, Sakakibara BM, Miller WC, Research Team SCIRE. Quality of life instruments and definitions in individuals with spinal cord injury: a systematic review. Spinal Cord 2010;48:438–50.
3. The World Health Organization Quality of Life Assessment (WHO-QOL): development and general psychometric properties. Soc Sci Med 1998;46:1569–85.
4. Charlifue S, Post MW, Biering-Sorensen F, et al. International Spinal Cord Injury Quality of Life Basic Data Set. Spinal Cord 2012;50:672–5.
5. Tate D, Forchheimer M. Review of cross-cultural issues related to quality of life after spinal cord injury. Top Spinal Cord Inj Rehabil 2014;20:181–90.
6. Wood-Dauphinee S, Exner G, Bostanci B, et al. Quality of life in patients with spinal cord injury—basic issues, assessment, and recommendations. Restor Neurol Neurosci 2002;20:135–49.
7. Oyakawa A, Castineira C, Greve J, Furlan J, Forchheimer MB, Tate D. Reliability of the Portuguese version of the International Quality of Life Spinal Cord Injury Data Set. In: Abstract book 52nd annual meeting International Spinal Cord Society; 2013.
8. Post MW, Adriaansen JJ, Charlifue S, Biering-Sorensen F, van Asbeck FW. Good validity of the international spinal cord injury quality of life basic data set. Spinal Cord 2016;54:314–8.
9. New PW, Tate DG, Forchheimer MB, D'Andréa Greve JM, Parashar D, Post MWM. Preliminary psychometric analyses of the International Spinal Cord Injury Quality of Life Basic Data Set. Spinal Cord 2019;57:789–95.
10. Nachtegaal J, van Langeveld SA, Slootman H, Post MWM. Dutch-Flemish Spinal Cord Society. Implementation of a standardized dataset for collecting information on patients with spinal cord injury. Top Spinal Cord Inj Rehabil 2018;24:133–40.
11. Constitution of the World Health Organization. Am J Public Health Nations Health 1946;36:1315–23.
12. Post M. Definitions of quality of life: what has happened and how to move on. Top Spinal Cord Inj Rehabil 2014;20:167–80.
13. Post MW, Charlifue S, Biering-Sorensen F, et al. Development of the International Spinal Cord Injury Activities and Participation Basic Data Set. Spinal Cord 2016;54:530–4.
14. Rohn EJ, Post MW, van der Wind A, et al. A cross-cultural mixed methods validation study of the Spinal Cord Injury Quality of Life Basic Dataset (SCI QoL-BDS). Spinal Cord 2022;60:177–86.
15. Post MWM, Forchheimer MB, Charlifue S, D'Andrea Greve JM, New PW, Tate DG. Reproducibility of the international spinal cord injury quality of life basic data set: an international psychometric study. Spinal Cord 2019;57:992–8.
16. 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.
17. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. Arch Phys Med Rehabil 1989;70:857–60.
18. Biering-Sorensen F, Alexander MS, Burns S, et al. Recommendations for translation and reliability testing of International Spinal Cord Injury Data Sets. Spinal Cord 2011;49:357–60.
19. Masters GN. A Rasch model for partial credit scoring. Psychometrika 1982;47:149–74.
20. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. 2nd ed. Mahwah, NJ; London: Lawrence Erlbaum; 2007.
21. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York; London: McGraw-Hill; 1994.
22. Linacre JM. Winsteps Rasch measurement computer program user's guide. Beaverton, OR:Winsteps.com; 2015.
23. Wright BD, Linacre JM, Gustafson JE, Martin-Löf P. Reasonable mean-square fit values. Rasch Meas Trans 1994;8:370–1.
24. Christensen KB, Makransky G, Horton M. Critical values for Yen's Q3: identification of local dependence in the Rasch model using residual correlations. Appl Psychol Meas 2017;41:178–94.
25. Christensen KB, Salzberger T. Ask the experts: Rasch vs. factor analysis. Rasch Meas Trans 2012;26:1373–4.
26. Hagquist C, Andrich D. Recent advances in analysis of differential item functioning in health research using the Rasch model. Health Qual Life Outcomes 2017;15:181.
27. Holland PW, Wainger H. Differential item functioning. Erlbaum; 1993.
28. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Series B Stat Methodol 1995;57:289–300.
29. Strobl C, Kopf J, Zeileis A. Rasch trees: a new method for detecting differential item functioning in the Rasch model. Psychometrika 2015;80:289–316.

30. Komboz B, Strobl C, Zeileis A. Tree-based global model tests for polytomous Rasch models. Educ Psychol Meas 2018;78:128–66.
31. Smith Jr. EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3:205–31.
32. Smith RM, Miao CY. Assessing unidimensionality for Rasch measurement editor. In: Wilson M, ed. Objective measurement: theory into practice, Greenwich: Ablex; 1994.
33. Core Team R. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
34. Mair P, Hatzinger R. Extended Rasch modeling: the eRm package for the application of IRT models in R. J Stat Softw 2007;20:1–20.
35. Mueller M. iarm: item analysis in Rasch models. R package version 0.4.1. 2020. Accessed Feb. 2022. https://cran.r-project.org/web/packages/iarm/iarm.pdf.
36. Boone WJ, Staver JR, Yale MS. Rasch analysis in the human sciences. Dordrecht: Springer; 2014.
37. Schwartz CE, Stucky B, Rivers CS, Noonan VK, Finkelstein JA, Network R. Quality of life and adaptation in people with spinal cord injury: response shift effects from 1 to 5 years postinjury. Arch Phys Med Rehabil 2018;99:1599–608.
38. Rohn EJ, Tate DG, Forchheimer M, DiPonio L. Contextualizing the lived experience of quality of life for persons with spinal cord injury: a mixed-methods application of the response shift model. J Spinal Cord Med 2019;42:469–77.