

RESEARCH

Open Access



Developing a shortened spine functional index (SFI-10) for patients with sub-acute/chronic spinal disorders: a cross-sectional study

Charles Philip Gabel^{1^}, Antonio Cuesta-Vargas², Almir Vieira Dibai-Filho³, Hamid Reza Mokhtarinia^{4*}, Markus Melloh⁵ and Agnieszka Bejer^{6,7}

Abstract

Background Brief whole-spine patient-reported outcome measures (PROMs) provide regional solutions and future directions for quantifying functional status, evidence, and effective interventions. The whole-spine regional Spine Functional Index (SFI-25) is used internationally in clinical and scientific contexts to assess general sub-acute/chronic spine populations. However, to improve structural validity and practicality a shortened version is recommended. This study developed a shortened-SFI from the determined optimal number of item questions that: correlated with criteria PROMs being highly with whole-spine, moderately with regional-spine, condition-specific and patient-specific, and moderately-low with general-health and pain; retained one-dimensional structural validity and high internal consistency; and improved practicality to reduce administrative burden.

Methods A cross-sectional study ($n = 505$, age = 18-87 yrs., average = 40.3 ± 10.1 yrs) of sub-acute/chronic spine physiotherapy outpatients from an international sample of convenience. Three shortened versions of the original SFI-25 were developed using 1) qualitative 'content-retention' methodology, 2) quantitative 'factorial' methodology, and 3) quantitative 'Rasch' methodology, with a fourth 'random' version produced as a comparative control. The clinimetric properties were established for structural validity with exploratory (EFA) and confirmatory (CFA) factorial analysis, and Rasch analysis. Criterion validity used the: whole-spine SFI-25 and Functional Rating Index (FRI); regional-spine Neck Disability Index (NDI), Oswestry Disability Index (ODI), and Roland Morris Questionnaire (RMQ), condition-specific Whiplash Disability Questionnaire (WDQ); and patient-specific functional scale (PSFS); and determined floor/ceiling effect. A post-hoc pooled international sub-acute/chronic spine sample ($n = 1433$, age = 18-91 yrs., average = 42.0 ± 15.7 yrs) clarified the findings and employed the general-health EuroQol-Index (EQ-5D), and 11-point Pain Numerical Rating Scale (P-NRS) criteria.

Results A 10-item SFI retained structural validity with optimal practicality requiring no computational aid. The SFI-10 concept-retention-version demonstrated preferred criterion validity with whole-spine criteria (SFI-25 = 0.967, FRI = 0.810) and exceeded cut-off minimums with regional-spine, condition-specific, and patient-specific measures. An unequivocal one-dimensional structure was determined. Internal consistency was satisfactory ($\alpha = 0.80$) with no floor/ceiling effect. Post-hoc analysis of the international sample confirmed these findings.

[^]Deceased.

*Correspondence:

Hamid Reza Mokhtarinia

hrmokhtarinia@yahoo.com; hr.mokhtarinia@uswr.ac.ir

Full list of author information is available at the end of the article



Conclusion The SFI-10 qualitative concept-retention version was preferred to quantitative factorial and Rasch versions, demonstrated structural and criterion validity, and preferred correlation with criteria measures. Further longitudinal research is required for reliability, error, and responsiveness, plus an examination of the practical characteristics of readability and administrative burden.

Keywords Spine, Musculoskeletal, Assessment, Patient-reported outcome measure, Functional limitation, Clinometric

Background

Functional status measurement is frequently determined with patient-reported outcome measures (PROMs) as they provide optimal practicality, statistical coherence, and structural-validity [1]. For patients with spine disorders, there has been a progressive shift toward ‘whole-spine’ PROMs that measure status as a continuous functional kinetic-chain [2]. These have included static-PROMs, the Extended Aberdeen Spine Pain Scale (EASPS) [3], Functional Rating Index (FRI) [4], Spine Functional Index (SFI-25) [5], and the Computer Adaptive Testing (CAT) assessed Patient-Reported Outcomes Measurement Information System (PROMIS) for Physical Function (PROMIS-PF) [6, 7]. This whole-spine approach has high clinical relevance as a single, practical, psychometrically accurate, whole-spine PROM provides clinicians, researchers, and patients with a reduced administrative burden as multiple PROMs are no longer required for different regions and conditions [3, 8, 9]. This directly reduces the key barriers to PROM adoption [10, 11], complies with why a PROM is chosen and used under the essential nine pragmatic requirements [12, 13], and provides the capacity for a consistent spine single-score, broadened data-pooling, meta-analysis [14], and the capacity to demonstrate whether specific health-care delivery is effective or not [15].

To balance the psychometrics, practicality, and cultural transferability, any whole-spine PROM must comply with the ‘Consensus-based Standards for the selection of Health status Measurement Instruments’ (COSMIN) standards [16]. The SFI-25 does this, being stringently developed and initially conference presented in 2004, with E-publication in 2013, with publication delays due to Journal submission processes and PhD by Portfolio requirements, with the official publication in 2019 affected by similar Journal-related delays [5]. This eventual peer validation permitted the inclusion of the SFI-25 within a whole-spine static-PROMs systematic review that considered the FRI and EASPS, where both had recognized concerns [8], but consequently did not include the PROMIS-PF. The FRI critiques were that it be used with caution till more robust high methodological quality studies are found to support its measurement properties [17], that it has item-construct deficiencies [18], and questionable ability to adequately represent

whole-spine problems [8]. The EASPS, with 28–35 questions over four pages, is recognized as cumbersome with questionable COSMIN compliance [8].

The SFI-25 has had seven published validation studies [19–25], with a further comparative validation study under submission [26] and was most recently used in a chronic neck pain study [27]. These cultural-adaptation studies not only adapted and validated the SFI-25 for their specific linguistic and population requirements, but also performed criterion validity with multiple whole-spine, spine-region, general, and condition-specific populations. In each case, the SFI-25 was found preferable to the criteria PROMs that included the Neck Disability Index (NDI) [28], Oswestry Disability Index (ODI) [29], Roland Morris Questionnaire (RMQ) [30], and Whiplash Disability Questionnaire (WDQ) [31]. Additionally, suitable correlation was demonstrated with the patient-specific function scale (PSFS) [5] and EuroQuol-Index (EQ-5D) [19, 20], but less so with an 11-point pain numerical rating scale (P-NRS) [19] and the SF-36 PF scale [26]. However, the SFI-25’s structural validity was not unanimous with a shortened version recommended in most studies to improve practicality and structural validity.

The PROMIS-PF, using ‘CAT’ in varied spine-specific populations [32, 33], captures similar information to static-legacy PROMs [34, 35] but with greater efficacy and accuracy [6, 7, 36]. However, many populations lack the computing and internet accessibility necessary for PROMIS-PF, which, coupled with patient settings and computer literacy, must be considered [37]. Additionally, though content validity is sufficient, evidence quality in adult populations is low-moderate, particularly for single body areas and conditions, and elderly minority populations [38]. Further, minimal spine studies incorporated PROMIS-PF for its outcome measurement use, with substantial variability in domain validity between PROMIS-PF and criteria static-PROM [39]. Consequently, there remains a place and need for a simple-to-use, accurate, and practical whole-spine static-PROM with low administrative burden [12, 40].

The advocated methodologies to shorten PROMs are two-fold, qualitative and quantitative. Qualitative approaches use expert committee consensus with the ‘concept-retention’ method advocated for being judgmental and retaining the original PROMs theoretical domains [41]. Quantitative

approaches use statistical methods, with ‘factorial’ and ‘Rasch’ the most common [1, 41]. This study aimed to: 1) develop a shortened-SFI for assessing spine functional status; 2) determine the correlation between the shortened-SFI and whole-spine criteria; 3) assess the correlation between the shortened-SFI and regional-spine, condition-specific, and patient-specific; criteria 4) investigate the correlation between the shortened-SFI and general-health and pain criteria; 5) ensure that the shortened-SFI retains the psychometric characteristics of one-dimensional structural validity, high internal consistency, and no floor/ceiling effect; and 6) enhance the practicality of the shortened-SFI to reduce administrative burden.

Accordingly, we hypothesized that: 1) the developed shortened-SFI will exhibit a high correlation with whole-spine criteria; 2) the correlation between the shortened-SFI and regional-spine, condition-specific, and patient-specific criteria will be moderate; 3) the correlation between the shortened-SFI and general-health and pain criteria will be moderate to low; 4) the psychometric properties of the shortened-SFI, including one-dimensional structural validity, high internal consistency, and absence of floor/ceiling effects, will be retained; and 5) practical enhancements made to the shortened-SFI will result in a reduction of administrative burden.

Methods

Study design

This cross-sectional study (n=505) was conducted to shorten the SFI-25 to the SFI-10. All subjects provided written informed consent with the study approved by the Ethical Committee of the Universidade Federal do Maranhão (approval protocol number 4.284.203).

Subjects

Participants were recruited from physiotherapy outpatients (n=505, age=18-87 yrs., av.=40.3 ± 10.1 yrs.,

female=50.5%, Table 1). There was no significant difference between the obtained SFI-10 scores by female (8.01 ± 6.14) and male (7.48 ± 5.60) (p=0.317). *Inclusion criteria* were a medical/allied-health practitioner referral with a spine musculoskeletal disorder (MSD) diagnosis, sub-acute/chronic symptoms ≥ 2 weeks, age ≥ 18 years, written language competence, and informed written consent. *Exclusion criteria* were pregnancy, age < 18 years, and red-flag signs [19, 23].

The post-hoc international sample (n=1433, age=18-91 yrs., av.=40.3 ± 10.1 yrs., female=58.4%, Table 1) included retrospective de-identified data obtained with permission from the original researchers of three additional published SFI-25 cross-cultural adaptation studies [19, 22, 23] and a further data set from a completed MSc research study [26] that has progressed to journal submission.

Measures

The spine functional index (25 items)

The SFI-25 has 25 item-questions with a 3-point response option ‘Yes’ (score=1), ‘Partly/Sometimes’ (score=1/2) and ‘No’ (score=0). Item-questions have a biopsychosocial 60:40 item-question ratio [5, 42] with 15 ‘General’ (#1–15) and 10 ‘Region-specific’ (#16–25) item-questions. ‘Raw Score’ (0–25) totals from the summation of all item responses. The final score (0–100%: 0% = ‘worst possible’; 100% = ‘normal’/‘preinjury function’) is calculated by: [100-(Raw Score × 4)] [5], with two missing responses permitted and substituted with the average score of all responded item-questions [5].

Functional rating index

The FRI has 10 item-questions with five short-descriptive response options (0–4 Likert visual NRS). ‘Raw Score’ (0–40) totals from the summation of all item responses. The final score, (0–100%: (0% = ‘no problem/pain’;

Table 1 Demographics for all study participants

Study		n	Age x	Age SD	Female n	Female%	Cx	Tx	Lx	Multi
SFI-10 development and validation (Australian in English Language)		505	40.1	13.0	255	50.5	149	48	350	53
Bejer 2019 [19]	Polish SFI	225	45.7	16	135	60	92	38	114	19
Tonga 2015 [23]	Turkish SFI	285	45	1.0	207	73	129		151	5
Mokhtarinia 2018 [22]	Persian SFI	224	38.8	10.9	104	46.4	112	13	87	12
Freitas 2023 [26]	Brazilian SFI	194	29.1	8.5	136	70.1	38	60	96	
							36.3%	11.1%	55.7%	6.2%
	TOTALS	1433	40.3 yrs	10.1 yrs	837	58.4%	520	159	798	89
			Range							
			18-87 yrs							

SFI indicates 25-item Spine Functional Index, n number, x mean, SD standard deviation, % percent, Cx cervical, Tx thoracic, Lx lumbar

* Subregion % values include multi-area individuals within each of their symptomatic regions making the total > 100%

100% = 'worst possible') is calculated by: [Raw Score \times 2.5] with one response permitted for substitution [4].

Each of the other spine-regional and general criteria PROMs are described in their original respective publications.

Development and psychometric assessment of the SFI-10

'Development' the shortened version of the SFI-25 was done through a-priori determination of the minimum number of item-questions necessary to retain structural validity and optimal practicality without a computational aid. The *minimum number* was guided by Spearman-Brown's 'k value' [43, 44], the *optimal number* by completion/scoring-time, accuracy, and no computational aid being required [12, 45]. Additionally, one-dimensional structural integrity was required along with face, content and criterion validity (Pearson's or Spearman's r), plus internal consistency (Cronbach's α : scale-level > 0.75 ; item-level > 0.65) [46, 47].

Four methodological approaches obtained the required optimal number of item-questions.

Version A: qualitative 'concept-retention' [41] obtained consensus agreement using the "Ishikawa" qualitative process [48] from semi-structured interviews with 'Expert' ($n=7$) and 'Patient' ($n=4$) focus-groups [49]. The 'Expert-group' was four males and three females, included three physiotherapists, an occupational therapist, orthopedic specialist, registered nurse, and biostatistician. The 'Patient-group', two males two female, paired for neck and back MSD.

Version B: quantitative 'factorial' used exploratory factorial analysis (EFA) with polychoric correlation matrix and robust diagonally weighted least squares (RDWLS) extraction (Factor loading > 0.40) [50] to obtain the highest loading items. Retained factors were defined through parallel analysis with random exchange of observed data and robust promin rotation [51, 52]. Model adequacy used Kaiser-Meyer-Olkin (KMO > 0.70) and Bartlett's sphericity tests ($p > 0.05$) from FACTOR software. The confirmatory factorial analysis (CFA) model used fit indices for: chi-square/degrees of freedom (chi-square/df < 3), root means square error of approximation (RMSEA < 0.08 ; CI = 90%), comparative fit index (CFI > 0.90), and Tucker-Lewis index (TLI > 0.90) from R-Studio software with Lavaan and SemPlot packages [53].

Version C: quantitative 'Rasch' extracted and confirmed the optimal-items through 'Person Abilities' and 'Item Difficulties' (preferred mean = 0.00); Personal separation reliability (PSR: cut-off > 0.70); one-dimensionality (Martin-Löf test: $p > 0.05$; $n = 800$ limit), and Principle Component Analysis (PCA) of Rasch-residuals Eigenvalues (cut-off = Linacre's value < 2.0) [54]; 'infit-outfit' statistic elimination (range: 0.5/0.7–1.3/1.5); item characteristic

curves (ICCs); and thresholds proximity (three-response options crossover, with item difficulties ordering); 'Wright-mapping' (for item spacing and redundancy) [55]; 'Algorithmic item-ranks' and 'Item-distances'; and Rasch corrected raw-scores (for person ability) [53].

Version D: 'Random' selected 10 random computer-generated items.

Validation' selected the optimal shortened-version as that with the highest criterion-correlation (Pearson's r) with whole-spine 'Gold Standard' criteria, the SFI-25 ($n = 505$, $r \geq 0.95$) and FRI ($n = 343$, $r \geq 0.70$) [47], supported by criterion validity cut-off scores ($r \geq 0.50$) with spine-regional instruments the NDI ($n = 143$), ODI ($n = 194$), RMQ ($n = 31$), and WDQ ($n = 70$), and the patient specific PSFS ($n = 174$). Full-sample structural validity was verified with EFA, CFA, and Rasch analysis, along with internal consistency (scale cut-off level: $\alpha \geq 0.75$, item level: $\alpha \geq 0.65$) and floor/ceiling effect from the percentage frequency for the highest/lowest scores (15% cut-off) [16].

A post-hoc pooled international sample ($n = 1433$) was analyzed to clarify structural validity, internal consistency, and floor/ceiling effect. Additionally, extracted Polish-study shortened-SFI scores ($n = 225$) [19] were compared with the SFI-25, spine-regional NDI ($n = 49$), ODI ($n = 86$), and general-health EQ-5D ($n = 125$) and pain P-NRS ($n = 225$); with the SFI-10 data referenced against the SFI-25. The Spearman r correlation coefficient (SCC) was used for non-normally distributed data.

The sociodemographic data and questionnaire scores used mean (\bar{x}) and standard deviation (SD) in SPSS version 17 with significance: $p < 0.05$. The Kolmogorov-Smirnov test verified data-distribution. Factorial/Rasch analyses were blinded to minimize bias.

Results

The 'Development' indicated the *minimum number* of item-questions was $n = 8$ (Spearman-Brown $k = 3.33$). The *optimal number* of item-questions was $n = 10$, (from options of SFI-8, 10, 12 and 15 items), as this required no computational aid and retained the biopsychosocial 60:40 item-question ratio with six 'General' (#1–6) and four 'Region-specific' (#7–10) items (Fig. 1). The item-reduction and selection process confirmed face and content validity. The SFI-10 'Raw Score' (0–10) is totaled from the summation of all item responses with the final score from: [100 - (Raw Score \times 10)], with one missing response and substitution permitted.

'Validation' selected the 10-item qualitative concept-retention version as it: provided the highest Pearson's r criterion-correlation with whole-spine criteria (SFI-25, $r = 0.967$, $n = 505$; FRI, $r = 0.810$, $n = 343$, Table 2); being supported by spine-regional and patient-specific

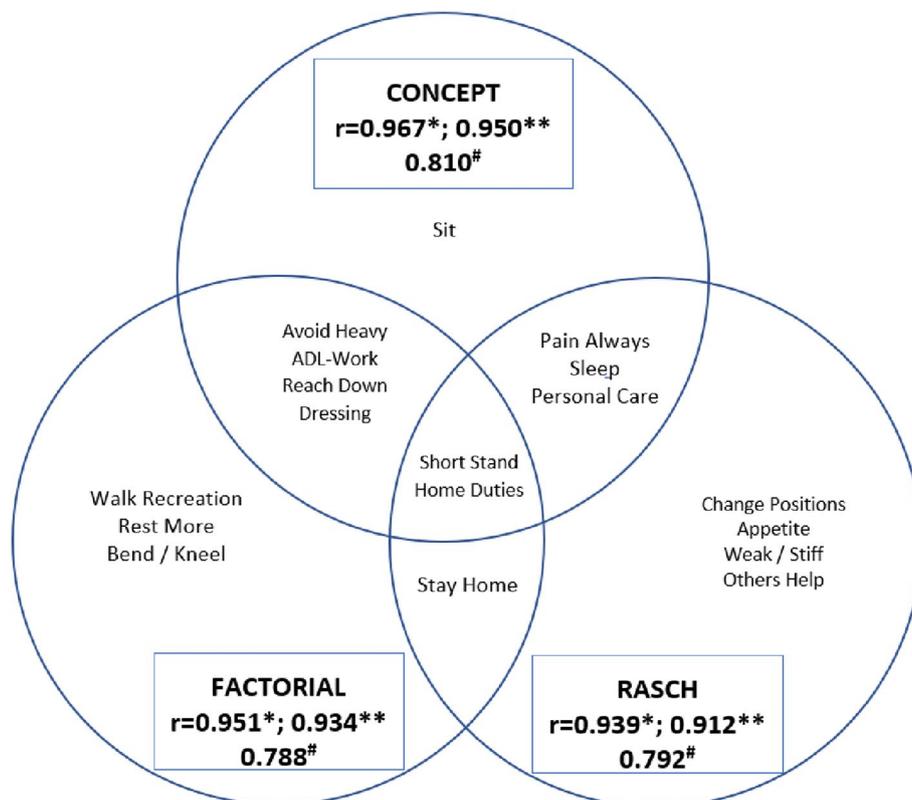


Fig. 1 Reduction Approaches: Items and overlap of the three SFI-10 reduction methods and Pearson’s r correlation with: original SFI-25 (*n=505, **n=1433); and # = FRI (n=343). Preferred SFI-10 was Concept version with the highest r value. Concept=qualitative concept-retention method; Factorial=factor analysis method; Rasch=Rasch analysis method. (Only two items were shared in all three methods. Concept shared the most items, then Factorial, then Rasch)

Table 2 Criterion validity comparing SFI-10 versions with the SFI-25 and FRI

Phase	SFI-25	SFI-10			
		Version A Qualitative 'Concept'	Version B Quantitative 'Factorial'	Version C Quantitative 'Rasch'	Version D Computerized 'Random'
Australian sample (PCC)	SFI-25 (n = 505)	# 0.967*	0.951*	0.939	0.955*
	FRI (n = 343)	0.818	# 0.810**	0.788**	0.792**
Pooled International sample (SCC) [19, 22, 23, 26]	SFI-25 (n = 1433)	# 0.950*	0.934	0.912	0.938

SFI indicates Spine Functional Index, PCC Pearson’s Correlation Coefficient for normally distributed data, FRI Functional Rating Index

*PCC/SCC: r ≥ 0.95 with the SFI-25; and **r > 0.70 with the FRI as the indicator of potential suitability to substitute for the SFI-25

PCC/SCC Highest value was the preferred version

criteria ($r \geq 0.70$, Table 3), except the NDI ($r = 0.693$) which approximated the $r = 0.70$ cut-off.

The ten items selected were: ‘Avoid Heavy Jobs,’ ‘Pain/Problem,’ ‘Duties/Chores,’ ‘Sleep,’ ‘Personal Care,’ ‘Daily Activity,’ ‘Dressing,’ ‘Sitting,’ ‘Standing,’ and ‘Reach/Bend Down.’

Structural validity met the a-priori requirements. The EFA identified a one-dimensional structure (Fig. 2) (KMO = 0.79; Bartlett’s test $p < 0.05$). The CFA confirmed EFA with fit indices: chi-square/df = 2.06, CFI = 0.952, TLI = 0.939, RMSEA (90% CI) = 0.073 (0.049, 0.096) (Table 4). Appropriate factor loadings

Table 3 Criterion validity for the SFI-25 and SFI-10 from existing published research

PROM	Australian (PCC)		Polish [19] (SCC)		Turkish [23] (PCC)	Persian [22] (PCC)	Spanish [20] (PCC)	Chinese [24] (PCC)	Korean [21] (PCC)	Greek [25] (PCC)
	SFI-25	SFI-10	SFI-25	SFI-10						
SFI-25										
		0.965 (n = 505)		0.943 (n = 225)						
		0.950 (n = 1433)								
FRI	0.832 (n = 343)	0.810 (n = 343)			0.52 (n = 285)			0.66 (n = 265)	0.57 (n = 60)	
ODI	0.818 (n = 194)	0.780 (n = 194)	0.800 (n = 86)	0.797 (n = 86)	0.71 (n = 151)			0.75 (n = 142)	0.75 (n = 42)	0.58 (n = 60)
RMDQ	0.976 (n = 31)	0.893 (n = 31)				0.69 (n = 114)			0.75 (n = 42)	0.70 (n = 60)
NDI	0.748 (n = 143)	0.693 (n = 183)	0.548 (n = 49)	0.321 (n = 49)	0.58 (n = 129)	0.57 (n = 124)		0.61 (n = 118)	0.53 (n = 26)	
WDQ	0.883 (n = 70)	0.837 (n = 70)								
PSFS	0.743 (n = 174)	0.734 (n = 174)								
EQ-Index			0.408 (n = 125)	0.532 (n = 125)			0.42 (n = 226)			
P-NRS (SCC)			-0.377 (n = 225)	-0.423 (n = 225)						

SFI indicates Spine Functional Index, PCC Pearson's Correlation Coefficient for normally distributed data, SCC Spearman's Correlation Coefficient for non-normally distributed data, FRI Functional Rating Index, ODI Oswestry Disability Index, RMDQ Roland Morris Disability Questionnaire, NDI Neck Disability Index, WDQ Whiplash Disability Questionnaire, PSFS Patient Specific Index, EQ-Index EuroQol Index, P-NRS Pain Numerical Rating Scale

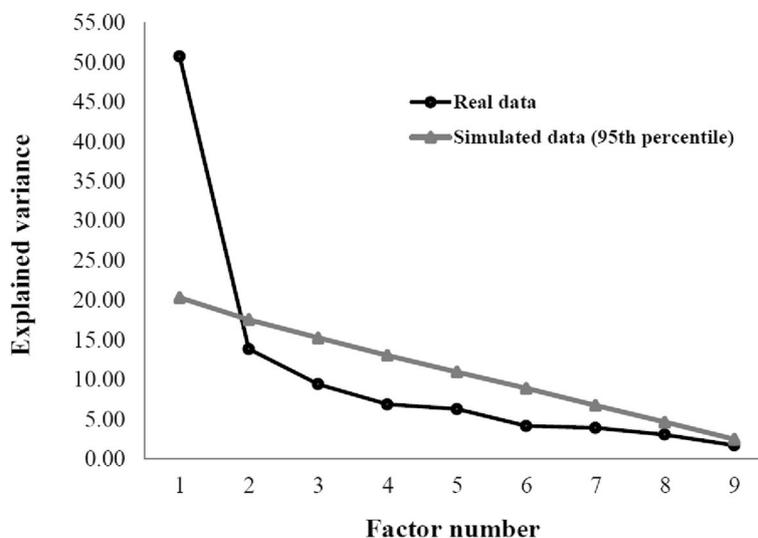


Fig. 2 Representative Scree Plot for SFI-10 EFA ($n = 505$). Post-hoc retrospective pooled samples ($n = 1433$) are similar with inflection at point #2

Table 4 Structural validity determination from factorial (CFA) analysis

Study group	Chi-square/df	CFI	TLI	RMSEA (90% CI)
$n = 505$	2.06	0.952	0.939	0.073 (0.049, 0.096)
$n = 1433$	2.92	0.961	0.950	0.069 (0.062, 0.077)

df indicates degrees of freedom, CFI comparative fit index, TLI Tucker-Lewis index, RMSEA root means square error of approximation, CI confidence interval

(> 0.40) were demonstrated between domains and items (Fig. 3).

Rasch analysis demonstrated adequate model fit (Table 5). ‘Person Abilities’ and ‘Item Difficulties’ indicated all tasks were within performance capacity, and PSR scores (0.71:0.79:0.75) exceeded the cut-off (> 0.70). One-dimensionality hypothesis (Martin-Löf test) was accepted ($p > 0.50$). Cut-off compliance was demonstrated for Rasch-residuals PCA (1.45–1.52:< 2.0), Infit-Outfit statistics (0.5–1.5), and item-difficulties (Table 5). Wright Map item-spacing and redundancy were acceptable, though some excess-spacing was present, but overall supported the selected item-shortening methodology. The ICC and Thresholds approximated a common point. Rasch corrected raw scores were completed (range:0–10). Rasch-analysis indicated the SFI-10 preserved the critical Rasch model-fit.

Internal consistency exceeded the a priori cut-off (scale level $\alpha = 0.803$, item level $\alpha \geq 0.65$). No floor/ceiling effects were found as minimum/maximum scores were < 15%.

Post-hoc analysis of the pooled international sample ($n = 1433$) confirmed the ‘concept-retention’ findings with the highest Pearson’s r criterion validity compared with

the whole-spine criteria (Table 2). The structural validity was one-dimensional where EFA used implementation of parallel analysis (KMO = 0.89, Bartlett’s test $p < 0.05$), and CFA fit indices approximated the main study (Table 4): chi-square/df = 2.92, CFI = 0.961, TLI = 0.950, RMSEA (90% CI = 0.069, 0.062, 0.077), with appropriate factor loadings (> 0.40) between domains and items. Rasch analysis approximated the main study and reinforced the one-dimensionality (Table 5). Internal consistency was high (scale level $\alpha = 0.863$, item level $\alpha > 0.65$) with no floor/ceiling effects.

The extracted Polish SFI-10 data criterion findings (Table 3) approximated the main study SFI-25 ($r = 0.943$ vs 0.965), ODI ($r = 0.797$ vs 0.780) except for the NDI ($r = 0.321$ vs 0.693). Similar correlations were found for the Polish SFI-25 with the spine-regional ODI, the EQ-5D and P-NRS criteria. The nine SFI-25 studies’ criteria findings were also comparable for the FRI, spine-regional, EQ-5D, and pain (Table 3).

Discussion

The study’s essential aims were achieved with a shortened SFI-10 developed. Face and concept validity were demonstrated by the reduction process with the criterion and structural validity confirmed by the psychometric analysis. The SFI-10 correlated highly with whole-spine criteria PROMs, moderately with region-specific, patient-specific, and condition-specific, and moderate-low for general-health and pain. Practicality was improved by 60%, though completion/scoring time/errors require quantification. The SFI-10 qualitative ‘concept-retention’ version demonstrated higher criterion validity with whole-spine criteria than the quantitative ‘factorial’ and

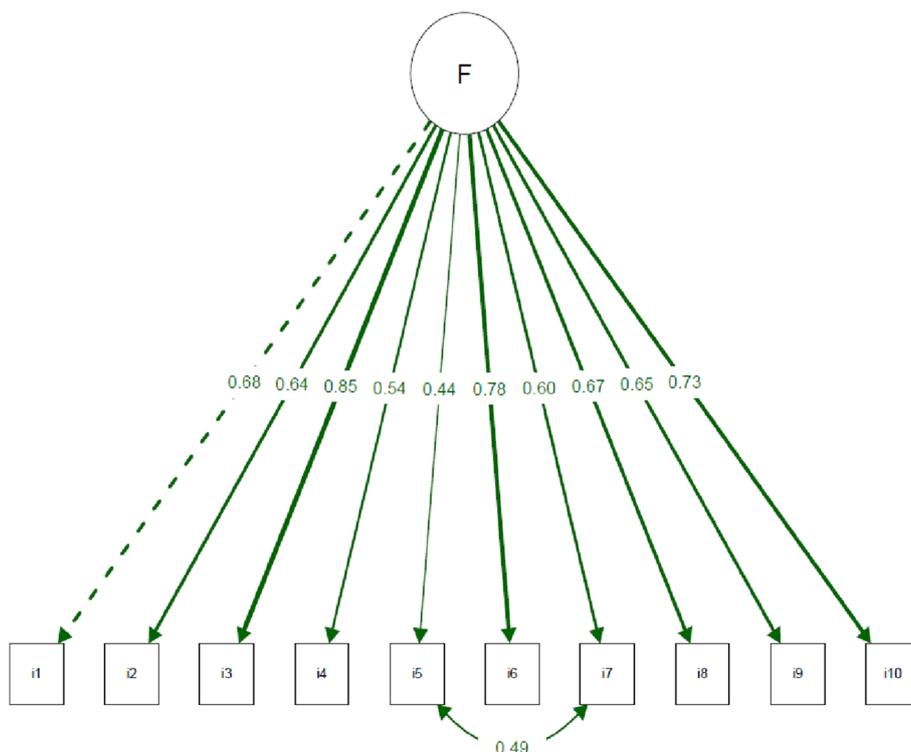


Fig. 3 Representative Scree Plot for SFI-10 EFA ($n = 505$). Post-hoc retrospective pooled samples ($n = 1433$) are similar with inflection at point #2

Table 5 Rasch analysis of the SFI-10 ($n = 505$ and $n = 1433$ are similar)

Item	SFI-25 Item #	Outfit		Infit		Item Difficulties	
		$n = 505$	$n = 1433$	$n = 505$	$n = 1433$	$n = 505$	$n = 1433$
1. Avoid Heavy Jobs	3.	1038	1167	0,967	1051	-0,675	-0,535
2. Pain / Problem	6.	1001	1038	0,991	1022	-0,833	-0,570
3. Duties / Chores	10.	0,648	0,699	0,715	0,771	-0,071	0,093
4. Sleep	11.	1236	1277	1171	1174	-0,463	-0,416
5. Personal Care	12.	1231	0,992	1105	1005	1414	0,841
6. Daily Activity	13.	0,785	0,805	0,812	0,822	-0,229	-0,056
7. Dressing	16.	1040	0,971	0,993	0,963	0,606	0,247
8. Sitting	20.	0,969	0,996	0,952	0,948	-0,308	-0,239
9. Stand	22.	0,985	0,938	1005	0,951	0,015	-0,058
10. Reach/Bend Down	24.	0,848	0,890	0,877	0,918	-0,196	-0,024

SFI indicates Spine Functional Index

‘Rasch’ versions, where both interestingly showed *lower PCC values* than the control/random (Table 2). Criterion validity was comparable with the FRI and slightly below the SFI-25 in the same sample and the original Australian SFI-25 study [5], but exceeded the Turkish [23], Korean [21] and Chinese [24] findings (Table 3).

Structural validity was unequivocally one-dimensional, being supported by factorial and Rasch analysis in the

full $n = 505$ sample and the post-hoc international sample ($n = 1433$). This complied with previous research recommendations that factor structure be improved as, although a dominant single-factor was present, 6–8 factors were demonstrated [5, 20, 22, 23]. Spine-regional and patient-specific criteria correlations approximated the SFI-25 findings, but the RMQ and NDI were notably lower (Table 3). However, SFI-10 spine-regional and

general-health criteria exceeded those of six SFI-25 studies [20–25] (Table 3).

Importantly, the SFI-10 retained the biopsychosocial 60:40 ratio conceptual model of general-versus-regional items [5, 42], which could not be maintained in the SFI-8, 12, and 15 item versions, each of which also required a computational aid. This biopsychosocial balance reduces risks of confounding ‘functional’ and ‘symptomatic’ change [56] while accommodating pain without potentially affecting responsiveness [57]. The increased SFI-10 practicality improved the scoring process without the need for a computational aid through a simple calculation of ‘ $\times 10$ ’ converting raw-scores to percentages [13, 45]. This should ensure lower administrative burden through reduced completion/scoring times [19, 40] and minimal potential errors [13], while complying with the essential nine pragmatic decisions for choosing and using a PROM [12]. In general, the popularity of short scales is explained by their need for reduced resources, particularly administrative burden and subsequent related costs [10, 40]. These findings reflect the two essential reasons for PROM shortening, practicality improvements and retaining validity and factor structure [16], as face, content, criterion, and structural validity must be retained [1, 46].

The preferred ‘concept-retention’ methodology supports similar PROM-shortening research where qualitative versions were superior to quantitative. This was demonstrated for the Quick-DASH (11-items) from the DASH (30-item) [41], though factor structure was not one-dimensional and practicality remained impaired as computational assistance was required. Similarly, concept-retention methodology produced the 10-item lower limb functional index (LLFI-10) from the LLFI-25 as a practical solution with one-dimensional validation in burns [58]. The 12-item Orebro Musculoskeletal *Screening* Questionnaire (OMSQ-12) improved the practicality of the original 21-item OMPainSQ and retained the critical psychometric characteristics for biopsychosocial risk screening [59, 60]. This contrasts with a qualitative ‘author-determined’ OMPainSQ-10 approach [61], where criterion validity was below the random version, as found in this study, and notably below the ‘concept-retention’ version [59]. The shortened NDI-5 combined qualitative and quantitative approaches, retained a one-dimensional structure [1, 56], and balanced psychometric and practical characteristics when compared to the 10-item version, the quantitative NDI-8 Rasch-version [57], and the NDI-7 factorial-version [1]. Various qualitative processes reduced the RMQ from 24 to 18 and 11 items [62], with the former, found preferable [62, 63]. However, no RMQ qualitative shortened version is available, and a computational aid remains necessary for all for practicality in

calculating the scores of all RMQ versions. However, the question remains as to what is ‘the optimal minimum number’ of item-questions that provides a sufficiently broad representation of the required domains [64], and can this be represented by only five items as per the NDI-5 [1, 56].

This study demonstrated and reinforced that a qualitative approach does produce a shortened-PROM that has balanced the requirements for critical psychometric characteristics and one-dimensional structural validity while concurrently improving practicality. Very short scales, below 10-items, increase the measurement error from lower precision [64], hence the SFI-10 version appears an appropriate solution. Consequently, this concept-retention qualitative item-reduction process can be confidently applied to similar regional PROMs to facilitate their application in clinical and research settings.

Study limitations and strengths

Study limitations include potential patient selection bias as recruitment was from primary contact and referred physiotherapy outpatients, consequently inpatient and community settings will need to be investigated. There is a lack of prospective data and repeated psychometric and practicality analysis. This leaves a knowledge gap in the test-retest reliability, responsiveness, and error scores, including both minimal detectable change and minimal clinically significant difference. Consequently, there is a need for longitudinal analysis, that includes patient-specific change, to clarify these psychometric properties. Further, the practical aspects of readability, missing responses, and administrative burden from completion and scoring times/errors must be quantified. Each of these latter limitations are now addressed in a subsequent study.

Study strengths included the large sample size and the clarification of findings in a further pooled international sample. Additionally, the SFI-10 development exceeded the minimal COSMIN standards and cut-off requirements. This incorporated the cross-sectional analysis and the pooled international sample from diverse populations with broad diagnoses.

Conclusions

This study developed a shortened 10-item SFI-10 whole-spine PROM and verified structural validity through factorial and Rasch analysis, criterion validity and internal consistency with no floor/ceiling effects. The pooled MSD population of diverse age, culture, and clinical settings supported potential generalizability for outpatient settings, but inpatient and community settings require investigation. The improved practicality and unequivocal one-dimensional factor structure

provided a summated score that is easily and rapidly determined without a computational aid. These attributes imply that the SFI-10 can be used in preference to the existing whole-spine and spine-regional PROMs in clinical and research settings. Further longitudinal research is currently underway to determine the critical psychometric characteristics of test-retest reliability, responsiveness, and error scores; and to quantify the practical characteristics of readability and administrative burden that include completion and scoring time/errors. Subsequently, a systematic review that includes the SFI-10 and published SFI-25 studies would further inform and clarify the clinimetric properties.

Abbreviations

AP	Parallel analysis
AUC	Area under the curve
CFA	Confirmatory factor analysis
CFI	Comparative fit index
COSMIN	COnsensus-based Standards for the selection of health status measurement instruments
Df	Degrees of freedom
EASPS	Extended aberdeen spine pain scale
EFA	Exploratory factor analysis
EQ-5D	EuroQol 5-dimensions questionnaire
ES	Effect size
FRI	Functional rating index
G-NRS	Global-numerical rating scales (perceived function)
ICC	Intra-class correlation coefficient
KMO	Kaiser-meyer-olkin test
K-S	Kolmogorov-smirnov statistic
MCID	Minimal clinically important difference
MDC	Minimum detectable change
MSD	Musculoskeletal disorder
N	Number
NDI	Neck disability index
ODI	Oswestry disability index
PCA	Principal component analysis
PCC	Pearson's correlation coefficient
P-NRS	Numerical rating scales (perceived pain)
PROM	Patient reported outcome measure
PSFS	Patient specific functional scale
PSR	Personal separation reliability
RDWLS	Robust diagonally weighted least squares
RMDQ	Roland morris disability questionnaire
RMSEA	Root means square error of approximation
ROC	Receiver operating curves
SCC	Spearman's correlation coefficient
SD	Standard deviation
sec	Seconds
SEM	Standard error of the measurement
SFI-10	Spine functional index, 10-items
SFI-25	Spine functional index, 25-items
SPSS	Statistical package for the social sciences
SRM	Standard response mean
TLI	Tucker-lewis index
WDQ	Whiplash disability questionnaire
α	Cronbach's alpha

Acknowledgements

Łukasz Deryło, MSc (Statistics), lukasz.derylo@gmail.com. Independent Statistician Cracow, Poland.
Henrique Yuji Takahasi, PT, hytakahasi@gmail.com - Postgraduate Program in Physical Education, Federal University of Maranhao, Sao Luis, Maranhao, Brazil.

Meihua Qian, PhD, mqian@g.clemson.edu - Department of Education and Human Development, College of Education, Statistics, Clemson University, SC, USA.

Ed Leahy, PT PhD, E.Leahy@latrobe.edu.au Department of Physiotherapy, La Trobe University, Australia.

Authors' contributions

All authors contributed to the study conception and design. Material preparation and data collection were performed by CPG, HR.M, A.C, A.B. Data analysis by CPG, AC, A.V, HR.M, and A.B with assistance from acknowledged contributors L.D, H.T and M.Q. the first original draft preparation was by CPG and M.M. Review and editing was performed by all authors who commented on subsequent versions of the manuscript and contributed to the final draft CPG, M.M, A.C, A.V, HR.M, and A.B. All authors read and approved the final manuscript.

Funding

No funding has been received by any agency in relation to this research.

Availability of data and materials

The data that support the findings of this study are available on request from the corresponding author, HR.M.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Universidade Federal do Maranhão (approval protocol number 4.284.203).

Written informed consent was obtained from all subjects. All methods of assessment were performed in accordance with the relevant guidelines and published papers.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Access Physiotherapy, PO Box 760, Coolumb Beach, Queensland 4573, Australia. ²Department of Psychiatry and Physiotherapy, Faculty of Medicine, Malaga University, Malaga, Spain. ³Federal University of Maranhao, Sao Luis, Maranhao, Brazil. ⁴Department of Ergonomics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. ⁵Queensland University of Technology, School of Public Health and Social Work, Brisbane, Australia. ⁶Institute of Health Sciences, Medical College, Rzeszow University, Rzeszow, Poland. ⁷The Holy Family Specialist Hospital, Rudna Mała 600, 36-060 Glogów Matopolski, Poland.

Received: 4 December 2023 Accepted: 12 March 2024

Published: 26 March 2024

References

1. Barreto FS, Avila MA, Pinheiro JS, Almeida MQG, Ferreira CSB, Fidelis-de-Paula-Gomes CA, Dibai-Filho AV. Less is more: five-item neck disability index to assess chronic neck pain patients in Brazil. *Spine (Phila Pa 1976)*. 2021;46(12):E688–93.
2. Hoffman J, Gabel CP. Expanding Panjabi's stability model to express movement: a theoretical model. *Med Hypotheses*. 2013;80(6):692–7.
3. Williams N, Wilkinson C, Russell IT. Extending the Aberdeen Back pain scale to include the whole spine: a set of outcome measures for the neck, upper and lower back. *Pain*. 2001;94(3):261–74.
4. Feise RJ, Menke JM. Functional rating index. A new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. *Spine*. 2001;26(1):78–86.
5. Gabel CP, Melloh M, Burkett B, Michener LA. The spine functional index: development and clinimetric validation of a new whole-spine functional outcome measure. *Spine J*. 2019;19(2):e19–27 Epub 2013 Oct 2025.

6. Papuga MO, Mesfin A, Molinari R, Rubery PT. Correlation of PROMIS physical function and pain CAT instruments with Oswestry disability index and neck disability index in spine patients. *Spine (Phila Pa 1976)*. 2016;41(14):1153–9.
7. Tishelman JC, Vasquez-Montes D, Jevotovsky DS, Stekas N, Moses MJ, Karia RJ, et al. Patient-reported outcomes measurement information system instruments: outperforming traditional quality of life measures in patients with back and neck pain. *J Neurosurg Spine*. 2019;1–6.
8. Leahy E, Davidson M, Benjamin D, Wajswelner H: Patient-Reported Outcome (PRO) Questionnaires for People With Pain in Any Spine Region. A Systematic Review. *Man ther* 2016, *Man Ther*(22):22–30.
9. Boody BS, Bhatt S, Mazmudar AS, Hsu WK, Rothrock NE, Patel AA. Validation of patient-reported outcomes measurement information system (PROMIS) computerized adaptive tests in cervical spine surgery. *J Neurosurg Spine*. 2018;28(3):268–79.
10. Crossnohere NL, Brundage M, Calvert MJ, King M, Reeve BB, Thorner E, Wu AW, Snyder C. International guidance on the selection of patient-reported outcome measures in clinical trials: a review. *Qual Life Res*. 2021;30:21–40.
11. Lam KC, Harrington KM, Cameron KL, Valier ARS. Use of patient-reported outcome measures in athletic training: common measures, selection considerations, and practical barriers. *J Athl Train*. 2019;54(4):449–58.
12. Kroenke K, Miksch TA, Spaulding AC, Mazza GL, DeStephano CC, Niazi SK, Illies AJC, Bydon M, Novotny PJ, Goyal A, et al. Choosing and using patient-reported outcome measures in clinical practice. *Arch Phys Med Rehabil*. 2022;103(5s):S108–s117.
13. Long C, Beres LK, Wu AW, Giladi AM: Patient-level barriers and facilitators to completion of patient-reported outcomes measures. *Qual Life Res* 2021, *E-Pub* 17 Sept 2021(1).
14. Morris T, Hee SW, Stallard N, Underwood M, Patel S. Can We Convert Between Outcome Measures of Disability for Chronic Low Back Pain? *Spine (Phila Pa 1976)*. 2015;40(10):734–9.
15. Hsiao CJ, Dymek C, Kim B, Russell B. Advancing the use of patient-reported outcomes in practice: understanding challenges, opportunities, and the potential of health information technology. *Qual Life Res*. 2019;28(6):1575–83.
16. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–57.
17. Bai Z, Shu T, Lu J, Niu W. Measurement properties of the functional rating index: a systematic review and Meta-analysis. *Spine (Phila Pa 1976)*. 2018;43(22):E1340–e1349.
18. Menke JM. The functional rating index: twenty years of invalid measurement. *Spine*. 2022;47(7):574–81.
19. Bejer A, Kupczyk M, Kwaśny J, Majkut A, Moskal K, Niemiec M, Gabel CP. Cross-cultural adaptation and validation of the polish version of the spine functional index. *Eur Spine J*. 2019;29(6):1424–34.
20. Cuesta-Vargas AI, Gabel CP. Cross-cultural adaptation, reliability and validity of the Spanish version of the spine functional index. *Health Qual Life Outcomes*. 2014;12(96).
21. In TS. The reliability and validity of the Korean version of the spine functional index. *J Phys Ther Sci*. 2017;29(6):1082–4.
22. Mokhtarinia HR, Hosseini A, Maleki-Ghahfarokhi A, Gabel CP, Zohrabi M. Cross-cultural adaptation, validity, and reliability of the Persian version of the spine functional index. *Health Qual Life Outcomes*. 2018;16(1):95.
23. Tonga E, Gabel CP, Karayazgan S, Cuesta-Vargas AI. Cross-cultural adaptation, reliability and validity of the Turkish version of the spine functional index. *Health Qual Life Outcomes*. 2015;13:30.
24. Zhou X, Xu X, Fan J, Wang F, Wu S, Zhang Z, Yang Y, Li M, Weil X. Cross-cultural validation of simplified Chinese version of spine functional index. *Health Qual Life Outcomes*. 2017;15:203.
25. Georgoudis G, Sotiropoulos S, Skouras A, Likourgia I, Retalis E, Karameri M. Musculoskeletal functional assessment: cross-cultural adaptation, validity, and reliability of the Greek version of the spine functional index (SFI). In: *World congress of physiotherapy*: 2019. Geneva; 2019.
26. FREITAS, Devyd Weyder do Nascimento. Tradução, Adaptação Transcultural E Validação Do Spine Functional Index Para O Português Brasileiro. 2023. https://bdtd.ibict.br/vufind/Record/JFMA_7383d9e4d10f84bc21094360e3a05b95.
27. Hessam M, Narimisa M, Monjezi S, Saadat M. Responsiveness and minimal clinically important changes to physical therapy interventions of Persian versions of copenhagen neck functional disability index, neck-bournemouth questionnaire and spine functional index questionnaires in people with chronic neck pain. *Physiother Theory Pract*. 2023;1–8. <https://doi.org/10.1080/09593985.2023.2210679>.
28. Vernon H, Mior S. The neck disability index: a study of reliability and validity. *J Manip Physiol Ther*. 1991;14(7):409–15.
29. Fairbank JCT, Pynsent PB. The Oswestry disability index. *Spine*. 2000;25(22):2940–52.
30. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low back pain. *Spine*. 1983;8:141–4.
31. Pinfold M, Niere KR, O'Leary EF, Hoving JL, Green S, Buchbinder R. Validity and internal consistency of a whiplash-specific disability measure. *Spine*. 2004;29(3):263–8.
32. Bernstein DN, Greenstein AS, D'Amore T, Mesfin A. Do PROMIS physical function, pain interference, and depression correlate to the Oswestry disability index and neck disability index in spine trauma patients? *Spine (Phila Pa 1976)*. 2020;45(11):764–9.
33. Passias PG, Pierce KE, Krol O, Williamson T, Naessig S, Ahmad W, Passfall L, Tretiakov P, Imbo B, Joujon-Roche R, et al. Health-related quality of life measures in adult spinal deformity: can we replace the SRS-22 with PROMIS? *Eur Spine J*. 2022;31(5):1184–8.
34. Hung M, Hon SD, Franklin JD, Kendall RW, Lawrence BD, Neese A, Cheng C, Brodke DS. Psychometric properties of the PROMIS physical function item bank in patients with spinal disorders. *Spine (Phila Pa 1976)*. 2014;39(2):158–63.
35. Brodke DS, Goz V, Voss MW, Lawrence BD, Spiker WR, Hung M. PROMIS PF CAT outperforms the ODI and SF-36 physical function domain in spine patients. *Spine (Phila Pa 1976)*. 2017;42(12):921–9.
36. Ziedas A, Abed V, Bench C, Rahman T, Makhni MC. Patient-reported outcomes measurement information system physical function instruments compare favorably to legacy patient-reported outcome measures in spine patients: a systematic review of the literature. *Spine J*. 2022;22(4):646–59.
37. Rafiq RB, Yount S, Jerousek S, Roth EJ, Cella D, Albert MV, Heinemann AW. Feasibility of PROMIS using computerized adaptive testing during inpatient rehabilitation. *J Patient-Rep Outcomes*. 2023;7(1):44.
38. Zonjee VJ, Abma IL, de Mooij MJ, van Schaik SM, Van den Berg-Vos RM, Roorda LD, Terwee CB. The patient-reported outcomes measurement information systems (PROMIS[®]) physical function and its derivative measures in adults: a systematic review of content validity. *Qual Life Res*. 2022;31(12):3317–30.
39. Haws BE, Khechen B, Bawa MS, Patel DV, Bawa HS, Bohl DD, Wiggins AB, Cardinal KL, Guntin JA, Singh K. The patient-reported outcomes measurement information system in spine surgery: a systematic review. *J Neurosurg Spine*. 2019;30(3):405–13.
40. Aiyegbusi OL, Roydhouse J, Rivera SC, Kamudoni P, Schache P, Wilson R, Stephens R, Calvert M. Key considerations to reduce or address respondent burden in patient-reported outcome (PRO) data collection. *Nat Commun*. 2022;13(1):6026.
41. Beaton DE, Wright JG, Katz JN, Group UEC. Development of the Quick-DASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am*. 2005;87(5):1038–46.
42. International Classification of Functioning, Disability and Health (ICF) www.who.int/icidh.
43. Spearman C. Correlation Calculated from Faulty Data. *Br J Psychol*. 1910;3:271–95. <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>.
44. Brown W. Some experimental results in the correlation of mental abilities 1. *Br J Psychol* 1904–1920. 1910;3(3):296–322.
45. Krosnick JA. *Questionnaire Design*. In: Vannette DL, Krosnick JA, editors. *The Palgrave handbook of survey research*. Cham: Springer International Publishing; 2018. p. 439–55.
46. de Vet HCW, Mokkink LB, Mosmuller DG, Terwee CB. Spearman-Brown prophecy formula and Cronbach's alpha: different faces of reliability and opportunities for new applications. *J Clin Epidemiol*. 2017;85:45–9.
47. Field A. *Discovering statistics using SPSS*. 3rd ed. London: SAGE Publications Ltd; 2009.
48. Ishikawa K, Loftus JH. *Introduction to quality control*. Tokyo: 3A Corporation; 1990.
49. Johnson RB, Onwuegbuzie AJ, Turner LA. Toward a definition of mixed methods research. *J Mixed Methods Res*. 2007;1(2):112.

50. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *MPR-online*. 2003;8(May):23–74.
51. Timmerman ME, Lorenzo-Seva U. Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychol Methods*. 2011;16:209–20.
52. Lorenzo-Seva U, Ferrando PJ. Robust Promin: a method for diagonally weighted factor rotation. *Lib Rev Peru Psicol*. 2019;25(1):99–106.
53. Mair P, Hatzinger R. Extended Rasch modeling: the eRm package for the application of IRT models in R. *J Stat Softw*. 2007;20(9):1–20.
54. Linacre M. Teaching Rasch measurement. *Rasch Meas Trans*. 2017;31(2):1630–1.
55. Wright BD, Masters GN. Rating scale analysis. Chicago, Illinois: MESA Press; 1982.
56. Walton DM, MacDermid JC. A brief 5-item version of the neck disability index shows good psychometric properties. *Health Qual Life Outcomes*. 2013;11(108).
57. van der Velde G, Beaton D, Hogg-Johnston S, Hurwitz E, Tennant A. Rasch analysis provides new insights into the measurement properties of the neck disability index. *Arthritis Rheum*. 2009;61(4):544–51.
58. Gittings PM, Heberlien N, Devenish N, Parker M, Phillips M, Wood FM, Edgar DW. The lower limb functional index - a reliable and valid functional outcome assessment in burns. *Burns*. 2016;pii: S0305-4179(16):30053–5.
59. Gabel CP, Melloh M, Yelland M, Burkett B. The shortened Örebro musculoskeletal screening questionnaire: evaluation in a work-injured population. *Man Ther*. 2013;18(5):378–85.
60. Shafeei A, Mokhtarinia HR, Maleki-Ghahfarokhi A, Piri L. Cross-cultural adaptation, validity, and reliability of the Persian version of the Örebro musculoskeletal pain screening questionnaire. *Asian Spine J*. 2017;11(4):520.
61. Linton SJ, Nicholas M, MacDonald S. Development of a short form of the Örebro musculoskeletal pain screening questionnaire. *Spine (Phila Pa 1976)*. 2011;36(22):1891–5.
62. Williams RM, Myers AM. Support for a shortened Roland-Morris disability questionnaire for patients with acute low back pain. *Physio Can*. 2001;53:60–6.
63. Macedo LG, Maher CG, Latimer J, Hancock MJ, Machado LA, McAuley JH. Responsiveness of the 24-, 18- and 11-item versions of the Roland Morris disability questionnaire. *Eur Spine J*. 2011;20(3):458–63.
64. Bowling A. Just one question: if one question works, why ask several? *J Epidemiol Community Health*. 2005;59(5):342–5.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.