Neurospine 2022;19(2):422-433.
https://doi.org/10.14245/ns.2143354.677

Neurospine
pISSN 2586-6583 eISSN 2586-6591

Original Article

Psychometric Properties of the Scoliosis Research Society Questionnaire (Version 22r) Domains Among Adults With Spinal Deformity: A Rasch Measurement Theory Analysis

Corresponding Author
Kati Kyrölä
https://orcid.org/0000-0002-4390-8735

Department of Orthopaedics, Central Finland Healthcare District, Hospital Nova, Hoitajantie 3, 40620 Jyväskylä, Finland
Email: kati.kyrola@ksshp.fi

Received: December 21, 2021
Revised: March 5, 2022
Accepted: March 10, 2022

Kati Kyrölä[1], Susanna Hiltunen[1,2], Mikko M. Uimonen[1], Jari Ylinen[3], Arja Häkkinen[3,4], Jussi P. Repo[1,5]

[1]Department of Surgery, Central Finland Healthcare District, Jyväskylä, Finland
[2]Department of Orthopaedics, Kuopio University Hospital, Kuopio, Finland
[3]Department of Physical Medicine and Rehabilitation, Central Finland Healthcare District, Jyväskylä, Finland
[4]Department of Health Sciences, University of Jyväskylä, Jyväskylä, Finland
[5]Department of Orthopaedics and Traumatology, Unit of Musculoskeletal Disease, Tampere University Hospital, Tampere, Finland

Objective: Adult spinal deformity (ASD) have lower health-related quality of life (HRQoL) compared to the general population. Applying Rasch measurement theory (RMT), this study tested the revised Scoliosis Research Society-22 (SRS-22r) HRQoL instrument among symptomatic adult patients with degenerative spinal disorders and varying degrees of ASD.
Methods: SRS-22r data from 637 outpatient spine clinic patients with degenerative spine conditions were investigated for unidimensionality, item/scale fit, differential item functioning (DIF), scale coverage/targeting, and person separation index (PSI) using RMT.
Results: Unidimensionality of the SRS-22r was not supported for either the total score or for 3 of its 5 domains. Item fit was acceptable for 11/22 items. The individual domains showed good coverage despite the degree of structural disorders. Ordered thresholds were achieved by merging response categories in some of the items. DIF towards age or sex was found in 11/22 items and in some domain items. The PSI exceeded 0.7 for the SRS-22r total score.
Conclusion: The individual domain scores of the SRS-22r perform better than the total score providing good coverage and targeting among patients with ASD. Refinements of items and domains may improve the structural validity of the instrument to meet the criteria for measuring ASD patients, even when multidimensionality persists.

Keywords: Spinal diseases, Rasch measurement theory, Latent trait theory, Revised Scoliosis Research Society-22, Outcomes research, Health-related quality of life

INTRODUCTION

Adult spinal deformity (ASD) is a common problem, with a reported prevalence of 32%, increasing with age.[1] In the population over age 60, a prevalence of degenerative scoliosis as high 68% has been reported.[2] The prevalence of spinal deformities is expected to rise further with increased life expectancy and population aging.[1] ASD patients have been reported to have lower health-related quality of life (HRQoL) when compared to the unaffected population standardized by age.[3] Patients with symptomatic spinal structural disorders have pain and limitations in functional abilities as well as problems with self-image and mental health issues.[3] Compared to other prevalent chronic diseases, such as arthritis, congestive heart failure, chronic lung

disease and diabetes, impaired HRQoL has been found to be even more prominent in patients with ASD.[4]

Spinal structural disorders in adults develop gradually over the years due to multiple etiologies, such as spinal degeneration, idiopathic scoliosis, neuromuscular or congenital origin, and obliquity of the pelvis.[5] Low back pain and sciatica are usually the main symptoms in the early phases of ASD.[6] Patients' HRQoL is affected in the early phases of sagittal malalignment long before the visible loss of sagittal or coronal balance.[7,8] Thus, it is essential to monitor the HRQoL of the patients with spinal disorders and detect problems associated with their spinal structural changes.

The Scoliosis Research Society questionnaire is a deformity-specific patient-reported outcome (PRO) instrument used to measure HRQoL outcomes of patients with spinal deformity.[3,8-10] Thus far the SRS questionnaire is the only disease-specific instrument available to measure HRQoL in patients of all ages with spinal deformity.

The revised Scoliosis Research Society-22 (SRS-22r) has 22 items[11,12] which are based on a 5-point symmetrical agree-disagree Likert scale. The response options are "very good, good, fair, poor, very poor", "none, mild, moderate, moderate to severe, severe," or "very often, often, sometimes, rarely, never." The items are scored from 1 to 5, with 1 being the worst and 5 the best result. The SRS questionnaire and the scoring guide are available free on the patient outcomes webpage: www.srs.org/professionals. The questionnaire can be scored as subtotals for individual domains (function/activity, pain, self-image/appearance, mental health) or as a total score (subtotal + satisfaction with management domain) (Table 1).

Since 1999 the SRS questionnaires have had different versions: 22, 22r (revised), 23, 24, and 30. In 2021, the SRS removed all but the revised version of SRS-22 from their webpage (www.srs.com) and recommends that all practitioners should utilize the SRS-22r, and its various translations. A translation code from all versions to the SRS-22r has been published by the developers of the questionnaires.[12] Thus far the SRS-22r has been more widely translated, validated, and revised among adolescent scoliosis population[11] rather than adults with degenerative spinal deformity.[3,13-15] Both the measurement properties of different translations[14,16] and structural validity of the SRS-22r total score[17-19] have been debated.

The SRS questionnaire has previously been found to be culturally and linguistically valid among adult patients with degenerative spinal complaints in Finland.[10] However, the structural validity of the SRS-22r domains has not been investigated using a partial credit model based on the Rasch measurement theory (RMT) model. Shortcomings in an instrument's structure may lead to bias when comparing large patient cohorts with different ages, diagnoses, cultures, and languages between centers or

Table 1. Domains, item numbers and content of the SRS-22r questionnaire domains

Domain	Items	Content of the domain
Function/activity	5, 9, 12, 15, 18	Level of ambulatory, professional and household activity
		Financial and social functioning
		Change in daily performance and leisure activity
Pain	1, 2, 8, 11, 17	Pain over the last 6 months, past month
		Pain at rest
		Level of pain medication
		Sick days due to pain
Self-image	4, 6, 10, 14, 19	Looks, outwards appearance
		Happiness with current body shape
		Attractivity among other people
		Personal relationships
Mental health	3, 7, 13, 16, 20	During last 6 months:
		Calm and peaceful or nervousness
		Downhearted and blue, feeling down
		Happiness
Satisfaction/dissatisfaction with management	21, 22	Are you satisfied with the results of your back management?
		Would you have the same maganement again if you had the same condition?

during follow-ups.[17] The RMT provides a tool to investigate the ability of the SRS-22r to measure a latent trait such as function, pain, self-image, mental health, or satisfaction with management. A scale measuring one latent trait can be considered unidimensional and linear, which is essential when measuring longitudinal changes in scores. Item and scale fit in the predefined model as well as construct validity can be tested using the RMT. Furthermore, response bias in each of the scale items can be tested for different degrees of structural disorders using differential item functioning (DIF). Currently, the RMT analysis can be considered one of the gold standard statistical techniques for instrument development and psychometric validation research.[20]

This study aimed to evaluate the applicability of the SRS-22r domains in clinical practice among all patients with subacute and chronic spinal degenerative conditions, with special emphasis on the level of structural disorders. Unidimensionality, item and scale fit, residual correlation, DIF, scale coverage/targeting, and person separation were investigated using RMT based on predefined hypotheses.

MATERIALS AND METHODS

A total of 991 consecutive patients with prolonged degenerative thoracolumbar disorder referred for specialist consultation to the spine clinic during 12 months in 2013 and 2014 were recruited to the study. Inclusion criteria were age over 18 years, ability to communicate in the official language and having full spine radiographs. Patients with specific health conditions, including malignancy, pregnancy, neuromuscular disease, or acute fracture were excluded. Altogether 874 patients met the inclusion criteria and 670 gave their written informed consent to participate in the study.

The patients completed the SRS-22r version of the questionnaire. Based on the spinal radiographs, the patients were classified into 3 categories of sagittal structural disorders severity (none or mild, moderate, and severe) according to the SRS-Schwab simplified classification as previously described.[8] All radiographic parameters were measured by a senior spine deformity surgeon. The study protocol was approved by the Central Finland Healthcare District Research Ethical Committee, Jyväskylä Finland (17U/2012).

RMT is a mathematical model designed to evaluate the properties of measurement instruments.[21] RMT analysis calculates the extent to which the observed responses fit the predefined measurement model responses and assesses the unidimensionality of the scale and the precision of measurement.[22,23] The model is based on latent trait theory and the application of additive conjoint measurement.[24] For Rasch analysis, sample sizes of ≥200 subjects can be considered very good with, sizes of 100–199 adequate, sizes of <50–100 doubtful, and sizes of <50 subjects inadequate.[25]

The study applied the RMT using Rumm2030 software to measure construct validity, model and individual item fit, and reliability. The polytomous partial credit model[26] was chosen. The RMT mathematic model describes the process and pertinent psychometric criteria for fit statistics and reliability.[26-29] Person estimation was conducted with the weighted maximum likelihood method. Analyses were conducted using statistical and illustrative tests in the software. Unidimensionality is one of the main assumptions of the RMT and refers to whether the items in a PRO instrument measure a single construct or a specific latent trait, such as pain or function.

The unidimensionality of the SRS-22r total score and its domains was examined. Principal component analysis (PCA) was conducted to define the "Rasch factor," i.e., the first factor identified with the highest eigenvalue. After identifying the Rasch factor, the existence of residual factors was examined by dividing the SRS-22r items into 2 groups according to their correlation coefficients with the second factor identified in the PCA. The items with correlation coefficients over +0.3 and those with correlation coefficients below -0.3 formed the 2 sets of items. The person estimates for each item for each patient were calculated in both sets of items. Conducting a series of independent-samples t-tests patient by patient, the estimates were compared between the 2 item sets. Less than 5% of significant t-tests at a 0.05 probability were used as the criterion for unidimensionality. Further, calculating residual correlations between each item pairs was used to identify item dependency. We used a value equal to or over 0.2 to recognize residual correlations.[30] Higher values are generally considered to indicate similarity between items and hence, either item redundancy or the existence of another latent trait after controlling for the influence of the primary factor. If unidimensionality was violated, testlets based on residual correlation between items were formed.

Testlest is formed by summing the response categories of suitable items into one item. Thus, testlets are item bundles that share a common content. To alleviate the influence of item dependency, each bundle is considered as a single polytomous item. The resulting polytomous RMT model is then applied to analyze the testlets. Items with residual correlations over 0.2 were combined to form testlests. Next, another set of independent-samples t-tests was conducted to investigate if the violation of

 https://doi.org/10.14245/ns.2143354.677

unidimensionality had been corrected. The authors hypothesized that the SRS-22r instrument and its 5 separate domains would exhibit a unidimensional structure.

To investigate the fit of the SRS-22r to the RMT, overall goodness-of-fit and item fit statistics were calculated. Chi-square (χ^2) values (item-trait interaction) and standardized fit residuals (item–person interaction) were investigated to identify item fit. χ^2-values can be used to investigate how well the difficulty in performing of the item meets the ability of the respondent and hence correctly discriminates between different states of the trait being measured. The standardized fit statistics provide information on how much a response differs from the model expectation. The outcome may be interpreted with very low standardized fit statistics indicating redundancy and high values indicating poor fit (Supplementary material 1). The authors hypothesized that the p-values of the chi-square test statistics after Bonferroni adjustment would be nonsignificant, indicating good fit.

The fit residual calculation (item–person interaction score) shows the level of divergence of the item for the persons who fit the model. The divergence calculation yields a residual score that estimates a standard normal distribution where the expected mean is 0 and standard deviation ± 1. In the RMT model, fit residuals between -2.5 and +2.5 are generally considered to indicate acceptable fit. Values below or over this range indicate over- or underdiscrimination in relation to average discrimination ability and thus poor fit of the item to the RMT model and measurement disturbances. A high residual fit can provide information on the redundancy of the given item, as the item may not contribute any new information to the scale.

The targeting ability of each item was examined by investigating the order of the thresholds of the response categories. A threshold indicates the point at which a 50% probability exists for the response to fall into either 1 of the 2 adjacent categories. Disordered thresholds indicate that the response categories resemble each other too closely to detect which category the answer should fall into. The authors hypothesized that the thresholds of the SRS-22r would be ordered.

The targeting and coverage of the SRS-22r scale were examined to investigate whether the questionnaire captures the whole spectrum of the subject matter in the sample as well as to obtain information about the range in which a questionnaire best functions in a distinct patient group. Person and item locations were then examined to determine whether the distribution of items matched the patient distribution on the scale. Differences in person-item distribution in subgroups by age, sex, and degree of spinal deformity were examined. Differences in person-item distribution and the functioning of the SRS-22r total score and of individual domains in the aforementioned subgroups were examined. Analysis of variance (ANOVA) was used to test the statistical significance between different groups. Differences in the mean score of those with different degrees of spinal structural disorders were tested. The authors hypothesized that no significant differences would be observed when the type I error rate (alpha) was set to 0.05. The person separation index (PSI) value was calculated to investigate the sensitivity of the instru-

Table 2. Patients' sociodemographic and clinical background details

Variable	Value
Age (yr)	54.8 ± 15.3
Female sex	358 ± 56.2
Body mass index (kg/m^2)	27.6 (4.8)
Marriage/live-in relation	452 (70.8)
Years of education	12 ± 3.6
Available for work	379 (59.4)
Smokers	153 (24.0)
Daily users of painkillers	338 (53.1)
Duration of current back pain (mo), median (IQR)	18 (7–60)
Diagnoses	
Scoliosis or kyphosis	34 (5.3)
Spondylolisthesis	98 (15.4)
Neural compression	307 (48.2)
Spondylosis without neural compression	198 (31.1)
Previous lumbar spine surgery	34 (5.3)
Severity of spinal structural disorders*	
None or mild	407 (64)
Moderate	159 (25)
Severe	71 (11)
SRS scores, maximum 5 points	
SRS total score	2,88 ± 0.56
SRS function/activity	2.82 ± 0.75
SRS pain	2.40 ± 0.76
SRS mental health	3.4 ± 0.88
SRS self-image/appearance	2.86 ± 0.67
SRS satisfaction with management	3.12 ± 0.75
SRS subtotal	2.88 ± 0.61

Values are presented as mean ± standard deviation or number (%) unless otherwise indicated.
IQR, interquartile range; SRS, Scoliosis Research Society.
*Spinal structural disorders were classified with the SRS-Schwab classification of adult spinal deformity.

ment to discriminate between patients of varying health status.[31] The PSI ranges between 0 and 1, with a higher value indicating better sensitivity. Values exceeding 0.7 are generally considered acceptable. The authors hypothesized a minimum PSI value of 0.80. DIF was used to test for possible response bias between subgroups in each item. DIF occurs when, for example, men and women within the same sample respond differently to an individual item. Uniform DIF means that the difference in probability remains constant at different levels of measurement. Nonuniform DIF, in turn, means that probabilities differ between groups at different levels of measurement. If the response distribution is similar between the subgroups under examination, then no DIF exists between the groups. If the distribution is similar in shape but follows different logit values, a uniform DIF (UD) is confirmed. If the shape of distribution is different, a nonuniform DIF (NUD) is confirmed. DIF was analyzed for age and sex. The authors hypothesized that there would not be DIF in any of the item towards age or sex. Bonferroni-adjusted ANOVA was used to identify potential item DIF.

RESULTS

A total of 637 patients with complete data and a signed informed consent (mean ± SD, aged 54.8 ± 15.3 years; 56.2% female) were included in the final analysis (effective response rate: 64.3%). Overall, 407 patients (64%) had none or mild, 159 moderate (25%) and 71 severe spinal structural disorders (11%). Patient characteristics are presented in Table 2.

1. SRS-22r Total Score

The unidimensionality of the SRS-22 items was not supported, as 20.57% of t-tests were significant at 0.05 probability (Table 3). A residual correlation over 0.2 was found between 38 item pairs (residual correlation matrix; see Supplementary material 2). Creating testlets using residual correlations or clinical and logically associations between relevant items did not produce unidimensional scale.

The item fit statistics calculated for each SRS-22r item revealed significant chi-square values after Bonferroni adjustment in 12 items (items 3, 7–8, 11, 13–14, 16–18, 20–22). Further, fit residuals falling outside the range of -2.5 to +2.5 were found in 11 items (items 7–8, 10–11, 13–17, 21–22).

When the 22 items of the SRS-22r were investigated as one scale, 16 of the 22 items had ordered thresholds. The remaining 6 items (11, 15, 17, 18, 19, and 22) had disordered thresholds.

Patients scored within the range set for the coverage of the scale. No statistically significant differences were observed for age ($p = 0.68$) or sex ($p = 0.06$) in the person and item distribution (Supplementary material 3). However, a moderate statistically significant difference ($p = 0.01$) was found for the person-item threshold distribution after grouping the patients by degree of structural spinal disorders but not after Bonferroni adjustment (Supplementary material 4). The PSI for the 22 items was 0.89 (Table 3). The PSI 0.89 indicated good degree to which patients can be differentiated into groups of person separation. Cronbach alpha was 0.89 for the SRS-22r. Five items (5, 6, 9, 10, 19) showed DIF for age and one item (12) DIF for sex (Table 4). As unidimensionality was not found for the total score with reasonable adjustments, the subsequent adjustment analyses are reported only for the domains of the SRS-22r.

2. Function/activity (F/A) domain

Table 3. Analysis monitoring statistics

SRS-22 score	Items		Persons		χ^2	df	p-value	PSI (extrems/no extrems)	Percentage of significant t-tests
	Location	Fit residual	Location	Fit residual					
Total score	0.0 ± 0.7	0.6 ± 3.6	-0.2 ± 0.8	-0.2 ± 1.4	918.7	198	0.000000	0.89/0.89	20.6
Function/activity	0.0 ± 0.9	0.7 ± 1.3	-0.3 ± 1.4	-0.3 ± 1.0	110.8	45	0.000000	0.77/0.77	4.9
Pain	0.0 ± 0.6	1.1 ± 1.3	-0.9 ± 1.1	-0.3 ± 1.0	67.3	45	0.02	0.67/0.60	6.6
Testlet 1	0.0 ± 0.9	-0.1 ± 4.7	-0.9 ± 1.2	-0.3 ± 1.0	57.0	45	0.11	0.85/0.85	1.4
Self-image	0.0 ± 0.9	0.3 ± 1.9	-0.2 ± 1.1	-0.3 ± 1.0	104.0	54	0.00005	0.76/0.76	11.1
Testlet 1	0.0 ± 1.5	-2.2 ± 7.8	-0.2 ± 1.3	-0.5 ± 0.8	16.5	18	0.56	0.76/0.77	3.1
Mental health	0.0 ± 0.7	-0.3 ± 2.0	-1.1 ± 2.6	-0.6 ± 1.1	49.1	45	0.31	0.90/0.89	6.3
Satisfaction with management	0.0 ± 1.2	-0.4 ± 0.3	-0.0 ± 1.2	-0.7 ± 1.1	11.0	11	0.45	0.33/0.23	1.9

Values are presented as mean ± standard deviation.
SRS-22, Scoliosis Research Society-22; df, degrees of freedom; PSI, person separating index.

https://doi.org/10.14245/ns.2143354.677

Table 4. Fit statistics and DIF for the items in the SRS-22r and its subscales

SRS-22r	Item	Missing	FR	χ^2	df	p-value	DIF age	DIF gender
Total score	1	0	0.09	9.84	9	0.364	-	-
	2	0	-0.37	9.04	9	0.4335	-	-
	3	1	-2.1	27.93	9	0.0009*	-	-
	4	0	0.14	7.80	9	0.5542	-	-
	5	3	-0.28	17.42	9	0.0426	UD*	-
	6	6	1.09	14.04	9	0.0426	UD*	-
	7	0	-3.12	28.40	9	0.0008*	-	-
	8	1	5.42	53.06	9	<0.0001*	-	-
	9	58	-1.19	10.95	9	0.2789	NUD*	-
	10	3	2.57	16.89	9	0.0504	UD*	-
	11	1	6.15	64.41	9	<0.0001*	-	-
	12	0	-1.98	23.76	9	0.0045	-	NUD*
	13	0	-2.87	50.69	9	<0.0001*	-	-
	14	0	-4.71	45.79	9	<0.0001*	-	-
	15	2	3.88	21.31	9	0.0113	-	-
	16	0	-3.9	50.74	9	<0.0001*	-	-
	17	204	5.09	122.73	9	<0.0001*	-	-
	18	5	-1.7	34.03	9	<0.0001*	-	-
	19	7	0.65	10.50	9	0.3118	UD*	-
	20	1	-2.47	40.76	9	<0.0001*	-	-
	21	9	3.33	31.98	9	0.0002*	-	-
	22	46	8.65	226.60	9	<0.0001*	-	-
Function	5	3	0.07	23.87	9	0.0045	UD*	-
	9	58	-1.15	17.16	9	0.0463	UD*	-
	12	0	0.04	18.29	9	0.0319	UD*	UD*
	15	2	2.15	18.74	9	0.0274	-	UD*
	18	5	1.79	32.23	9	<0.0002	-	-
Pain	1	0	-0.07	18.23	9	0.0062	-	-
	2	0	0.01	18.14	9	0.0095	-	-
	8	1	3.17	12.60	9	0.2561	-	-
	11	1	1.23	6.87	9	0.5856	-	-
	17	204	0.96	11.48	9	0.5541	-	-
Self-image	4	0	2.15	26.07	9	0.0019*	UD*	-
	6	6	-0.38	9.91	9	0.3577	UD*	-
	10	3	0.85	12.71	9	0.1761	-	NUD*
	14	0	-0.4	14.81	9	0.0964	-	-
	19	7	-2.47	26.54	9	0.0016	UD*	-
Mental health	3	1	1.04	3.652	9	0.9328	-	-
	7	0	-1.84	18.44	9	0.0305	-	-
	13	0	-0.71	9.95	9	0.3543	-	-
	16	0	-2.41	9.49	9	0.3932	-	UD*
	20	1	2.45	7.57	9	0.5778	NUD*	-
Satisfaction with management	21	9	0.58	4.23	9	0.6451	-	-
	22	46	0.21	6.73	9	0.2417	-	-

DIF, differential item functioning; SRS-22r, revised Scoliosis Research Society-22; FR, fit residual; df, degrees of freedom; UD, uniform DIF; NUD, nonuniform.
*Probabilities below Bonferroni adjustment.

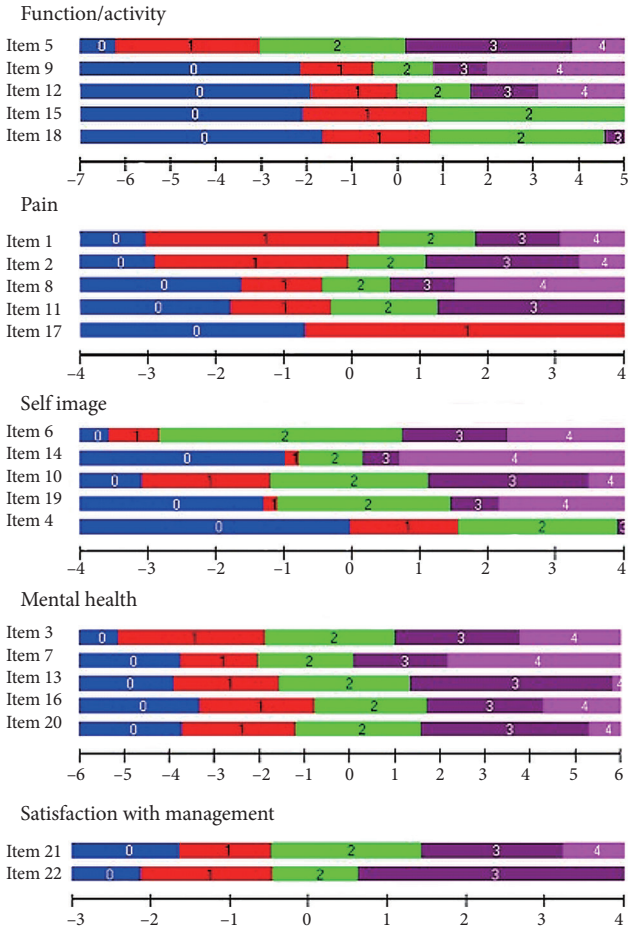


Fig. 1. Item response category thresholds (IRCTs) of the function/activity domain after merging response categories 1, 2, and 3 in item 1 and 2 and 3 in item 5. IRCTs of the pain domain after merging response categories 1 and 2 in item 4 and 0–2 and 3–4 in item 17. IRCTs of the self-image domain after merging response categories 1 and 2 in item 4 and 0–2 and 3–4 in item 17. IRCTs of the satisfaction with management domain after merging response categories 3 'probably not' and 4 'definitely not' in item 2. No merging to achieve ordered thresholds was required for the mental health domain.

In the function/activity domain, 4.9% of the t-tests were significant ($p < 0.05$), thereby supporting its unidimensionality (Table 3). Residual correlations over 0.2 were noted in 5 of the 10 item pairs (residual correlation matrix shown in Supplementary material 2). The item fit statistics in the function domain indicated good fit of the items to the RMT model (Table 3). The PSI for the domain was 0.77 and thus was below the hypothesized value of 0.8 (Table 3). Items 15 and 18 had disordered thresholds. Merging item response categories that score 1 to 3 in item 15 and 3 and 4 in item 18 led to ordered response category thresholds in each of the 5 function domain items (Fig. 1). The person-item threshold distribution showed only minor exceptions in the coverage of the function domain in the lower extremity of the scale (Supplementary material 5A). Subgroup analysis revealed significant differences between the severity classes in the person-item distribution of the Function domain, with higher severity patients having lower mean logit values ($p < 0.001$) (Fig. 2A). Uniform DIF for age and/or sex was observed in all the function domain items except item 18 (Table 4).

3. Pain Domain

In the original version of the pain domain, 6.6% of the t-tests were significant, indicating violation of the unidimensionality assumption (Table 3). Rescoring the items did not lead to a unidimensional scale structure, as the percentage of significant t-tests was unchanged. Nine out of 10 item pairs showed residual correlations over 0.2 (residual correlation matrix Supplementary material 2). The formation of a testlet by combining items 1 ('Which one of the following best describes the amount of pain you have experienced during the past 6 months?'), 2 ('Which one of the following best describes the amount of pain you have experienced over the last month?'), and 17 ('In the last 3 months have you taken any days off of work, including household work, or school because of back pain?') according to their residual correlations and contents led to a unidimensional scale, as 1.2% of the t-tests were significant (Table 3). All the pain domain items except item 8 showed acceptable fit residuals whereas, according to the Bonferroni-corrected chi-square tests, the item-trait interactions showed no significant distortions (Table 3). The PSI for the pain domain was 0.67 (Table 2). After testlet formation, the PSI increased to 0.85 (Table 3). In the pain domain, items 11 and 17 had disordered thresholds. Merging item response categories "nonnarcotics daily or less" in item 11 and categories "0–2 days absence" and "over 3 days absence" in item 17 led to ordered thresholds (Fig. 1). Overall, the patients scores indicated that coverage of the pain domain was good (Supplementary material 5B). The patients' logit values did not differ between the deformity severity subgroups ($p = 0.9$) (Fig. 2B). Age ($p = 0.0018$) was associated with the item location distribution. No DIF was observed in any of the pain domain items (Table 4).

4. Self-Image Domain

In the self-image domain, 9.91% of the t-tests were significant, indicating violation of the unidimensionality assumption (Table 3). A residual correlation of over 0.2 was found in 7 out of 10 item pairs. To achieve unidimensionality, items 4, 6, 10,

https://doi.org/10.14245/ns.2143354.677

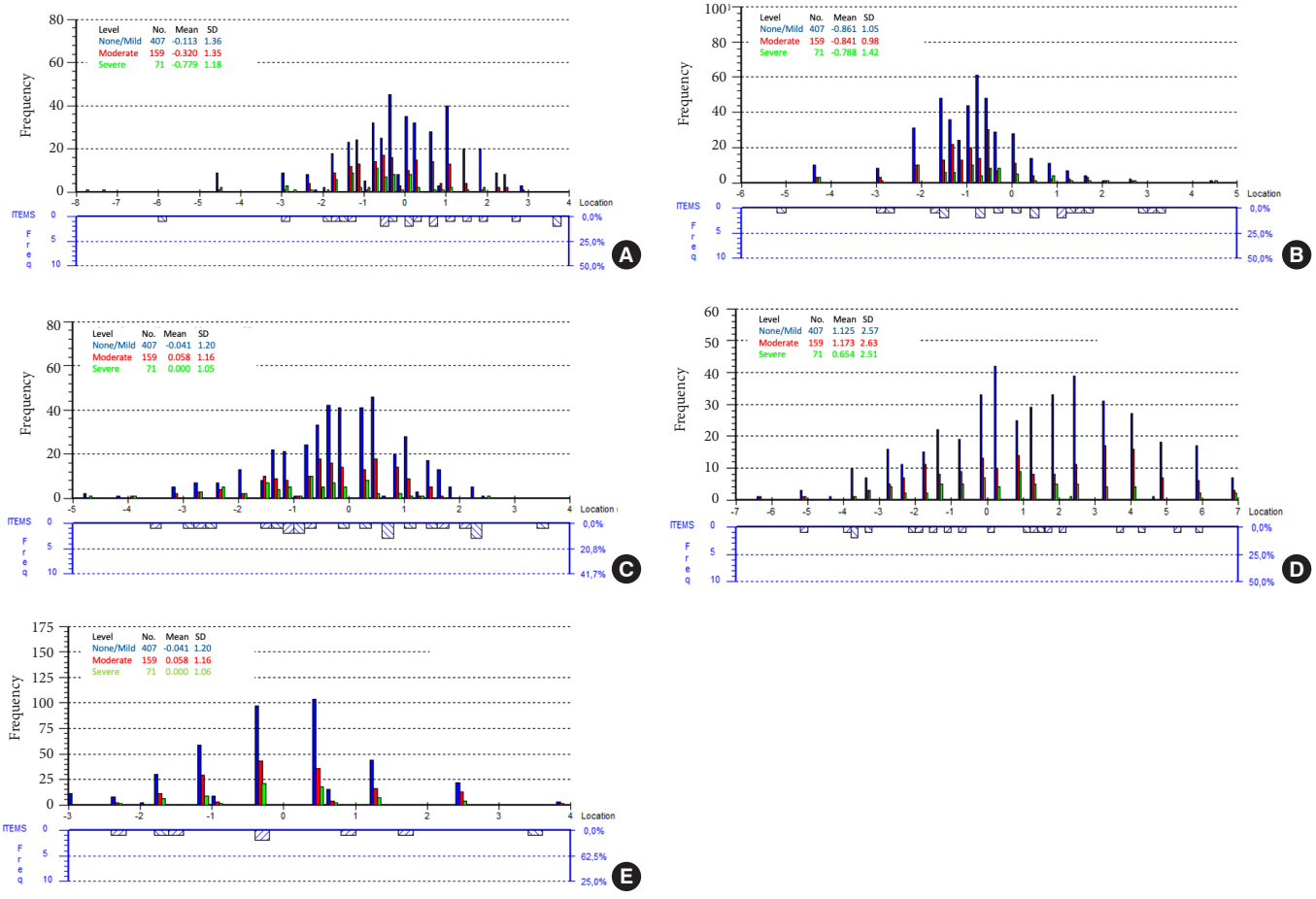


Fig. 2. Illustrations showing the person-item threshold distribution difference between groups of degree of spinal deformity in the distribution of person scores and items of the SRS-22r. Person-Item threshold distribution and degree of deformity. Mild or moderate deformity (blue), moderate (red), and marked structural disorder (green). Function/activity (A), pain (B), self-image (C), mental health (D), satisfaction with management (E). SRS-22r, revised Scoliosis Research Society-22; SD, standard deviation.

14, and 19 were pooled to form a testlet based on item content. The testlet reduced the proportion of significant t-tests to 3.1% (Table 3). All the self-image items except item 4 showed good fit to the RMT model (Table 3). Both the fit residual and Bonferroni-corrected chi-square statistic for item 4 indicated poor fit to the RMT model (Table 3). The PSI of the self-image domain was 0.76 (Table 3). To achieve ordered thresholds, response categories "somewhat happy" and "neither happy nor unhappy" in item 4 were merged (Fig. 1). Coverage of the self-image domain was good with minor discrepancy as 4 patients scored beyond the range of which the scale provided (Supplementary material 5C). There was a statistically significant difference in person-item distribution in age ($p < 0.001$), sex ($p = 0.01$), and degree of deformity ($p < 0.001$). Uniform DIF was observed across the age groups in items 4, 6, and 19, and nonuniform DIF was observed between sexes in item 10 (Table 4).

5. Mental Health Domain

In the mental health domain, the proportion of significant t-tests was 6.3%, and hence the domain was not unidimensional (Table 3). Residual correlations over 0.2 were found in 6 of the 10 item pairs. No clear testlet solution that would achieve unidensionality was available. All the items in the domain showed ordered thresholds as well as good fit to the RMT model (Fig. 1, Table 3). The PSI of the domain was 0.90. The domain covered the patients well, as only a few outliers were found at both extremities of the range (Supplementary material 5D). Coverage was equal in terms of degree of spine deformity ($p = 0.32$), age ($p = 0.64$), or sex ($p = 0.70$). DIF was detected in 2 out of 5 items (Table 4).

6. Satisfaction With Management Domain

The satisfaction with management domain met the criterion

Table 5. Results of the SRS-22r total score and it's domains

Reference value	Item fit	Ordered thresholds	Item DIF
	Nonsignificant after Bonferroni-correction		Nonsignificant after Bonferroni-correction
Total score	11/22	16/22	16/22
Function	4/5	3/5	1/5
Pain	5/5	3/5	5/5
Self-image	4/5	4/5	4/5
Mental well-being	5/5	5/5	3/6
Satisfaction with management	2/2	1/2	2/2

SRS-22r, revised Scoliosis Research Society-22; DIF, differential item functioning.

for unidimensionality, as 1.9% of the t-tests were significant (Table 3). No residual correlation was found between the 2 items. The item fit statistics indicated good fit to the RMT model of the 2 items (Table 3). Item 22 had disordered threshold categories. Merging response categories 'probably not' and 'definitely not' in item 22 produced ordered thresholds. The PSI value of the treatment satisfaction domain was 0.33 (Table 3). The patients' satisfaction with management scores showed that the domain covered the sample well (Supplementary material 5E). There was no discrepancy in the person-item distribution for age ($p = 0.21$), sex ($p = 0.26$), or degree of deformity ($p = 0.66$). No DIF was observed in either item (Table 4).

An overall summary of the RMT statistics for the SRS-22r domains is presented in Table 5.

DISCUSSION

The performance and structural validity of the SRS-22r questionnaire differed according to whether it was analyzed as the total score or as the individual domains. The SRS-22r total score showed poor structural validity when inserted into the RMT model. It seems that the construct validity of the SRS-22r improves when it is divided into distinct subscales. Nonetheless, the total score and its 5 domains provided sufficient coverage and targeting in all the spinal deformity severity categories.

In the present study on adults with degenerative spine conditions, the unidimensionality of the SRS-22r total score was not supported. The present findings are in line with previous findings of multidimensionality of the SRS questionnaires.[17,32] Jain et al.[18] and Caronni et al.[19] introduced a reduced, unidimensional and linear 7-item version (SRS-7) of the SRS-22 that met the Rasch criteria among adolescents with scoliosis. Jain et al.[33] validated the SRS-7 version on adults but the fit for RMT was not separately tested. Four of the pooled items in the SRS-7 were from the self-image and one each from the pain, function/activity, and mental health domains. The short version of the SRS instrument was found to be a good for assessing global changes but lacking the individual aspects of spinal deformity.[17,33] Mannion et al.[17] performed structural factor analysis on different linguistic versions of the SRS-22. They suggested that removing the worst fitting items (3, 14, 15, 17), one from each nonmanagement domains, would improve the multidimensional instrument together with standardization and validation of the items across language versions.[17] In the early revisions, items 17 and 18 were rephrased for the present SRS-22r after further adaptations among adolescents and adults.[11,13]

The concept of HRQoL is multidimensional,[34] and thus it is plausible that RMT analysis does not support the unidimensionality of the SRS questionnaires. Our findings indicate that the structural validity of the SRS-22r could be enhanced by re-evaluating its content and removing the afore mentioned potentially mis-fitting items. Moreover, the fact that the individual SRS-22r domains showed better structural validity leads us to recommend that the individual domain scores rather than the total score are used in clinical work and research. This might provide more accurate patient-reported outcome measure (PROM) data.

To the best of the authors' knowledge, the performance of no SRS instrument has previously been evaluated in participants with different degrees of ASD severity. The domains of the SRS-22r seem to work well irrespective of the degree of spinal structural disorders. The sample used in the present analysis presents the population visiting an orthopedic spine center due to prolonged degenerative thoracolumbar disorders. In adolescents, the SRS-22 was found to be inferior to the specific Spinal Appearance Questionnaire (SAQ) in detecting patients who required surgery and had greater curve magnitude.[35] In ASD, pain, disability, and sagittal structural changes cause deterioration in HRQoL and are the main drivers for seeking surgical treatment[36] instead of the deformity magnitude. Adults also comprise a very heterogenous group of people as to their spinal disorders and medical conditions compared to adolescents with idiopathic scoliosis. The degree of spinal deformity may affect the completion of the total score so that more respondents have for example higher scores from harder items affecting the person-item distribution. Further studies could focus on performing the RMT analysis for different stages of spinal deformity, a task that

https://doi.org/10.14245/ns.2143354.677

was beyond the scope of this study. Also, the previously studied HRQoL instruments failed to account sufficiently for neurogenic injury or impairment.[37] The SRS version 30 total score has been structurally evaluated in relation to radicular symptoms,[9] but further studies are required to evaluate the validity of the SRS-22r for measuring neurogenic impairment. Structural validity of the SRS or other deformity-specific HRQoL instruments has not been mathematically analyzed in large patient cohorts or with RMT. The SRS-22r domains are reported separately in several studies, but to the author's best knowledge few studies report results in comparison between the individual SRS domains.[14,38]

Compared to the SRS-22r total score or other domains, Function/activity most optimally met the RMT model criteria. It was found to be unidimensional with both good item fit and coverage and an acceptable PSI level. The domain achieved ordered thresholds in all items after merging the response categories in items asking about current level of activity and the frequency of going out compared to friends. Potential item response bias, compromising fit to the RMT model, was noted when patients were divided into subgroups by age or sex. Majority of the function/activity domain items showed DIF with at least one tested age or sex group. Patients with high degree of sagittal deformity also had lower logit values and hence a different person-item distribution in the function/activity domain. This may indicate that physical functioning and the capability to perform and daily activities depend on the degree of spinal deformity and that this difference is detected by the SRS-22r Function/activity domain.[3,4]

The pain domain showed good item fit and sensitivity, and no DIF was found. The domain items ask about pain during the last 6 months, during the past month and during rest, the use of pain medication, and the frequency of absence from work or school (item 17). The last 2 items may also measure other traits that patients cannot clearly differentiate from their spinal condition when filling in the questionnaire. Item 17 showed misfit to the RMT rating scale structure parameters (Andrich thresholds), indicating that the response categories did not match the item's intended meaning. Altogether 40% of the participants were not available for employment or school, which could explain the confusion over the response categories in this older population. Adapting item 17 to better serve ASD patients who may be students, in employment or retired, can be recommended. Pain was the only SRS-22r domain that showed no response bias between the age or sex groups. The domain functioned well across all degrees of spinal deformity. However, the pain scale differed between the age groups in its coverage and targeting.

Although the self-image/appearance domain did not show unidimensionality, it showed good item fit in 4/5 of the items, sensitivity and coverage. Item 4 ("If you had to spend the rest of your life with your back shape as it is right now, how would you feel about it?") showed potential misfit to the RMT model. The sensitivity of the domain was acceptable. It was also multidimensional; however, removing or modifying item 4 might improve the fit of the domain to the RMT model. This domain might also improve the value of the SRS instrument in measuring HRQoL in all degrees of ASD, as the other spine questionnaires do not place similar emphasis on the emotional and psychological functions.[39]

The mental health domain items were taken with permission from the Rand Corporation's SF-36 instrument. All the mental health items are good measures of mental well-being problems, as demonstrated by their ordered thresholds, good sensitivity, coverage and fit to the RMT model. However, in this study, the mental health domain was not unidimensional. In another patient cohort with prolonged back pain and associated depression and distress, the SRS-22r has also shown a multidimensional structure.[40] Potential age-related response bias was found for item 20, which asks how often the respondent has been a happy person, and sex-related bias for item 16, which asks whether the respondent has felt downhearted and blue. Such bias may be explained by the multidimensionality of the measured trait and respondents' interpretation of the positive vs negative tone of the item (happy vs. blue).

Satisfaction with management is rarely covered in the PROMs used for spinal problems. This 2-item unidimensional domain showed good coverage and fit to the RMT model and no DIF. Merging 2 response categories (probably not and definitely not) in item 22 resulted in ordered thresholds. The domain is simple and short, has good structural and psychometric validity, and can be recommended in clinical use.

The strength of this study was the consecutive-sample cohort of symptomatic adult patients with a wide range of different degrees of spinal deformities. The dropout rate of the recruitment was low, and thus the result can be generalized to real-life studies of this patient population. RMT was applied in a sufficient sample size to provide reliable information on the psychometric and structural properties of the SRS-22r. Furthermore, to our best knowledge, the individual SRS-22r domains have not been evaluated with the RM among adults. Chi-square statistics can be sensitive to large sample sizes. As our sample size was ample, it could potentially result in significant chi-square statistics, even for a well-fitting measure. The limitations of the current study

are that analysis only included mostly preoperative patients and that the study was as a single-center study conducted in one spine clinic. Furthermore, the SRS-22r and its domains scores could be structurally investigated and developed among adult patients who have undergone surgery due to spinal deformity.

CONCLUSION

The results of the present RMT analysis show that, among ASD patients, the individual domain scores of the SRS-22r perform better than the total score. Refining items and domains may improve the validity of the instrument for use with adult patients with spinal deformities, even when multidimensionality between domains persists. The questionnaire largely performed equally across age and sexes. The SRS-22r domains were able to differentiate between degrees of spinal deformity.

SUPPLEMENTARY MATERIALS

Supplementary materials 1-5 can be found via https://doi.org/10.14245/ns.2143354.677.

NOTES

Conflict of Interest: The authors have nothing to disclose.

Funding/Support: This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution: Conceptualization: KK, JY, AH, JR; Data curation: KK, MU, JR; Formal analysis: MU, JR; Methodology: KK, JY, AH, JR; Project administration: KK, JY, AH; Visualization: KK; Writing - original draft: KK, SH, JR; Writing - review & editing: KK, SH, MU, JY, AH, JR.

ORCID

Kati Kyrölä: 0000-0002-4390-8735
Susanna Hiltunen: 0000-0001-5962-5302
Mikko M. Uimonen: 0000-0001-6609-1345
Jari Ylinen: 0000-0002-5701-4265
Arja Häkkinen: 0000-0002-5779-1259
Jussi P. Repo: 0000-0002-4688-5698

REFERENCES

1. Ames CP, Scheer JK, Lafage V, et al. Adult spinal deformity: epidemiology, health impact, evaluation, and management. Spine Deform 2016;4:310-22.
2. Schwab F, Dubey A, Gamez L, et al. Adult scoliosis: prevalence, SF-36, and nutritional parameters in an elderly volunteer population. Spine (Phila Pa 1976) 2005;30:1082-5.
3. Berven S, Deviren V, Demir-Deviren S, et al. Studies in the Modified Scoliosis Research Society outcomes instrument in adults: validation, reliability, and discriminatory capacity. Spine (Phila Pa 1976) 2003;28:2164-9; discussion 2169.
4. Pellisé F, Vila-Casademunt A, Ferrer M, et al. Impact on health related quality of life of adult spinal deformity (ASD) compared with other chronic conditions. Eur Spine J 2015;24:3-11.
5. Aebi M. The adult scoliosis. Eur Spine J 2005;14:925-48.
6. Mendoza-Lattes S, Ries Z, Gao Y, et al. Natural history of spinopelvic alignment differs from symptomatic deformity of the spine. Spine (Phila Pa 1976) 2010;35:E792-8.
7. Buckland AJ, Vira S, Oren JH, et al. When is compensation for lumbar spinal stenosis a clinical sagittal plane deformity? Spine J 2016;16:971-81.
8. Kyrölä K, Repo J, Mecklin JP, et al. Spinopelvic changes based on the simplified SRS-Schwab adult spinal deformity classification: relationships with disability and health-related quality of life in adult patients with prolonged degenerative spinal disorders. Spine (Phila Pa 1976) 2018;43:497-502.
9. Kyrölä K, Häkkinen AH, Ylinen J, et al. Further validation of the Scoliosis Research Society (SRS-30) questionnaire among adult patients with degenerative spinal disorder. Disabil Rehabil 2021;43:98-103.
10. Kyrölä K, Järvenpää S, Ylinen J, et al. Reliability and validity study of the finnish adaptation of Scoliosis Research Society questionnaire version SRS-30. Spine (Phila Pa 1976) 2017;42:943-9.
11. Asher MA, Lai SM, Glattes RC, et al. Refinement of the SRS-22 health-related quality of life questionnaire function domain. Spine (Phila Pa 1976) 2006;31:593-7.
12. Lai SM, Burton DC, Asher MA, et al. Converting SRS-24, SRS-23, and SRS-22 to SRS-22r: establishing conversion equations using regression modeling. Spine (Phila Pa 1976) 2011;36:E1525-33.
13. Bridwell KH, Cats-Baril W, Harrast J, et al. The validity of the SRS-22 instrument in an adult spinal deformity population compared with the Oswestry and SF-12: a study of response distribution, concurrent validity, internal consistency, and reliability. Spine (Phila Pa 1976) 2005;30:455-61.
14. Arima H, Carreon LY, Glassman SD, et al. Cultural variations in the minimum clinically important difference thresholds

for SRS-22R after surgery for adult spinal deformity. Spine Deform 2019;7:627-32.
15. Mannion AF, Impellizzeri FM, Leunig M, et al. EUROSPINE 2017 FULL PAPER AWARD: Time to remove our rose-tinted spectacles: a candid appraisal of the relative success of surgery in over 4500 patients with degenerative disorders of the lumbar spine, hip or knee. Eur Spine J 2018;27:778-88.
16. Monticone M, Nava C, Leggero V, et al. Measurement properties of translated versions of the Scoliosis Research Society-22 Patient Questionnaire, SRS-22: a systematic review. Qual Life Res 2015;24:1981-98.
17. Mannion AF, Elfering A, Bago J, et al. Factor analysis of the SRS-22 outcome assessment instrument in patients with adult spinal deformity. Eur Spine J 2018;27:685-99.
18. Jain A, Sponseller PD, Negrini S, et al. SRS-7: A valid, responsive, linear, and unidimensional functional outcome measure for operatively treated patients with AIS. Spine (Phila Pa 1976) 2015;40:650-5.
19. Caronni A, Zaina F, Negrini S. Improving the measurement of health-related quality of life in adolescent with idiopathic scoliosis: the SRS-7, a Rasch-developed short form of the SRS-22 questionnaire. Res Dev Disabil 2014;35:784-99.
20. Cano S, Barrett L, Zajicek J, et al. Beyond the reach of traditional analyses: using Rasch to evaluate the DASH in people with multiple sclerosis. Mult Scler 2011;17:214-22.
21. Pallant J, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 2007;46:1-18.
22. Andrich D. Rating scales and Rasch measurement. Expert Rev Pharmacoecon Outcomes Res 2011;11:571-85.
23. Bond TG, Yan Z, Heene M, et al. Applying the Rasch model. 4th ed. Fourth edition. New York (NY): Routledge; 2020.
24. Tennant A, McKenna SP, Hagell P, et al. Application of Rasch analysis in the development and application of quality of life instruments. Value Health 2004;7 Suppl 1:S22-6.
25. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10:22.
26. Andrich D. Rasch models for measurement. Newbury Park (CA): Sage Publications; 2002.
27. 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.
28. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. J Rehabil Med 2003; 35:105-15.
29. Hobart JC, Cano SJ, Zajicek JP, et al. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol 2007;6:1094-105.
30. 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.
31. Wright B, Stone M. Reliability in Rasch measurement. In: Wright BD, Stone MH. Measurement essentials. 2nd ed. Wilmington (DE): Wide Range Inc.; 1994. p. 157-65.
32. Rothenfluh DA, Neubauer G, Klasen J, et al. Analysis of internal construct validity of the SRS-24 questionnaire. Eur Spine J 2012;21:1590-5.
33. Jain A, Lafage V, Kelly MP, et al. Validity, reliability, and responsiveness of SRS-7 as an outcomes assessment instrument for operatively treated patients with adult spinal deformity. Spine (Phila Pa 1976) 2016;41:1463-8.
34. Sajid MS, Tonsi A, Baig MK, et al. Health-related quality of life measurement. Int J Health Care Qual Assur 2008;21:365-73.
35. Bernstein DN, Papuga MO, Sanders JO, et al. Evaluating the correlation and performance of PROMIS to SRS questionnaires in adult and pediatric spinal deformity patients. Spine Deform 2019;7:118-24.
36. Diebo BG, Varghese JJ, Lafage R, et al. Sagittal alignment of the spine: what do you need to know? Clin Neurol Neurosurg 2015;139:295-301.
37. Stienen MN, Maldaner N, Sosnova M, et al. Lower extremity motor deficits are underappreciated in patient-reported outcome measures: added value of objective outcome measures. Neurospine 2020;17:270-80.
38. Bourghli A, Boissière L, Kieser D, et al. Multiple-rod constructs do not reduce pseudarthrosis and rod fracture after pedicle subtraction osteotomy for adult spinal deformity correction but improve quality of life. Neurospine 2021;18:816-23.
39. Repo JP, Ponkilainen VT, Häkkinen AH, et al. Assessment of construct validity of the Oswestry Disability Index and the Scoliosis Research Society-30 Questionnaire (SRS-30) in patients with degenerative spinal disease. Spine Deform 2019; 7:929-36.
40. Leino P, Magni G. Depressive and distress symptoms as predictors of low back pain, neck-shoulder pain, and other musculoskeletal morbidity: a 10-year follow-up of metal industry employees. Pain 1993;53:89-94.

Supplementary material 1. Item-trait interaction formulae in Rasch measurement theory.

Müller and Kreiner describe the formulae behind the RUMM2030 analysis as follows: "The item-trait interaction is calculated using group residuals for item chi-square fit statistics as the sum of squared group residuals." Persons are grouped into classes (g) depending on the individual scores (Formula 1).

$$X_i^2 = \sum_{g} Z_{gi}^2 = \sum_{g} \left[\frac{\sum_{v \in g} X_{vi} - \sum_{v \in g} E(X_{vi})}{\sqrt{\sum_{v \in g} Var(X_{vi})}} \right]^2$$

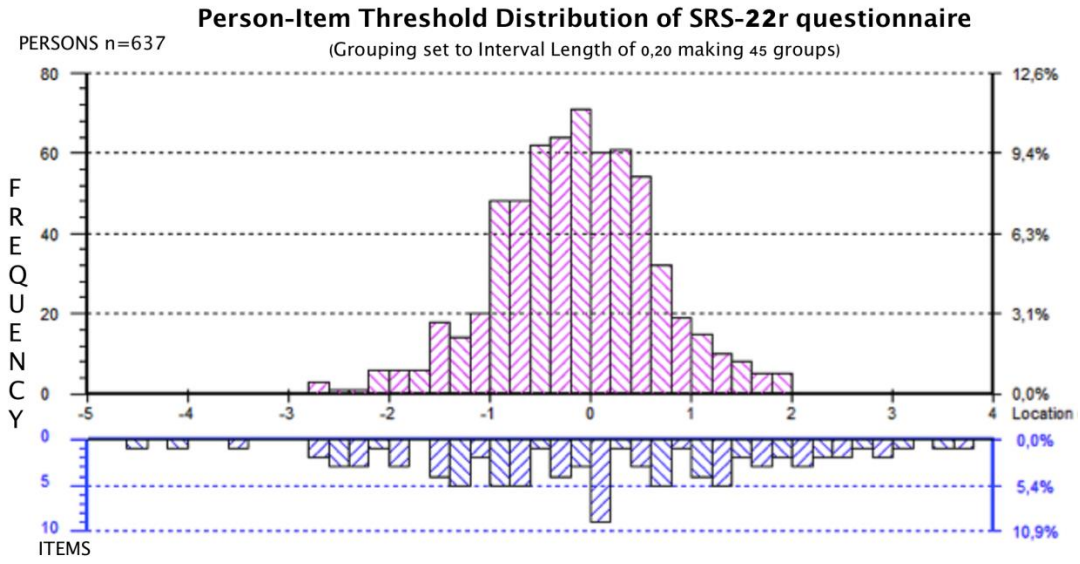
The total "item-trait interaction" chi-square test statistic is the sum of item chi-square test statistic, as shown below in Formula 2.

$$X^2 = \sum_{i} X^2{}_i \, , df = k \cdot df_i$$

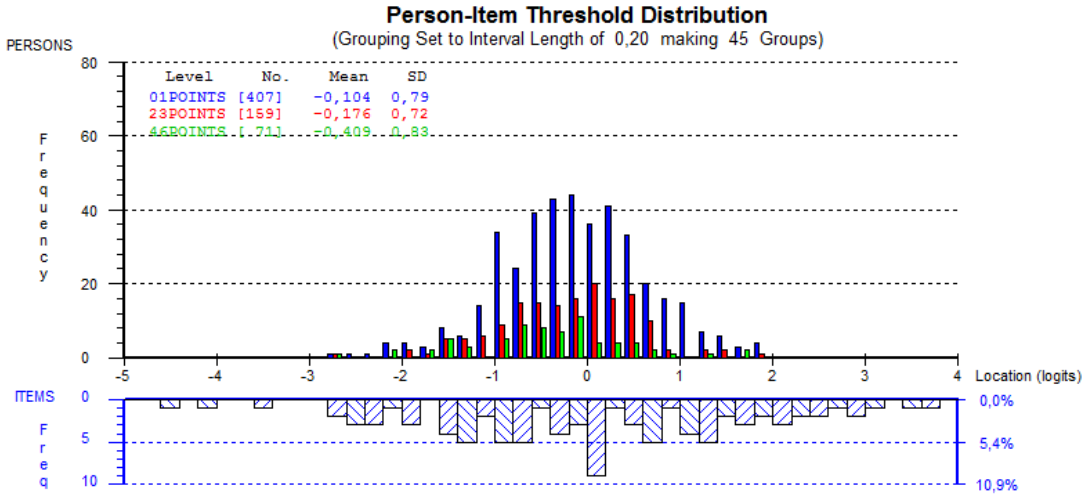
Adapted from: Müller M, Kreiner S. Item fit statistics in common software for rasch analysis. Copenhagen; 2015. Available from: https://ifsv.sund.ku.dk/biostat/annualreport/images/2/2f/Research_Report_15-06.pdf.

Supplementary material 2. Residual correlation matrix SRS-22 domains. SRS-22r, revised Scoliosis Research Society-22.

https://doi.org/10.14245/ns.2143354.677

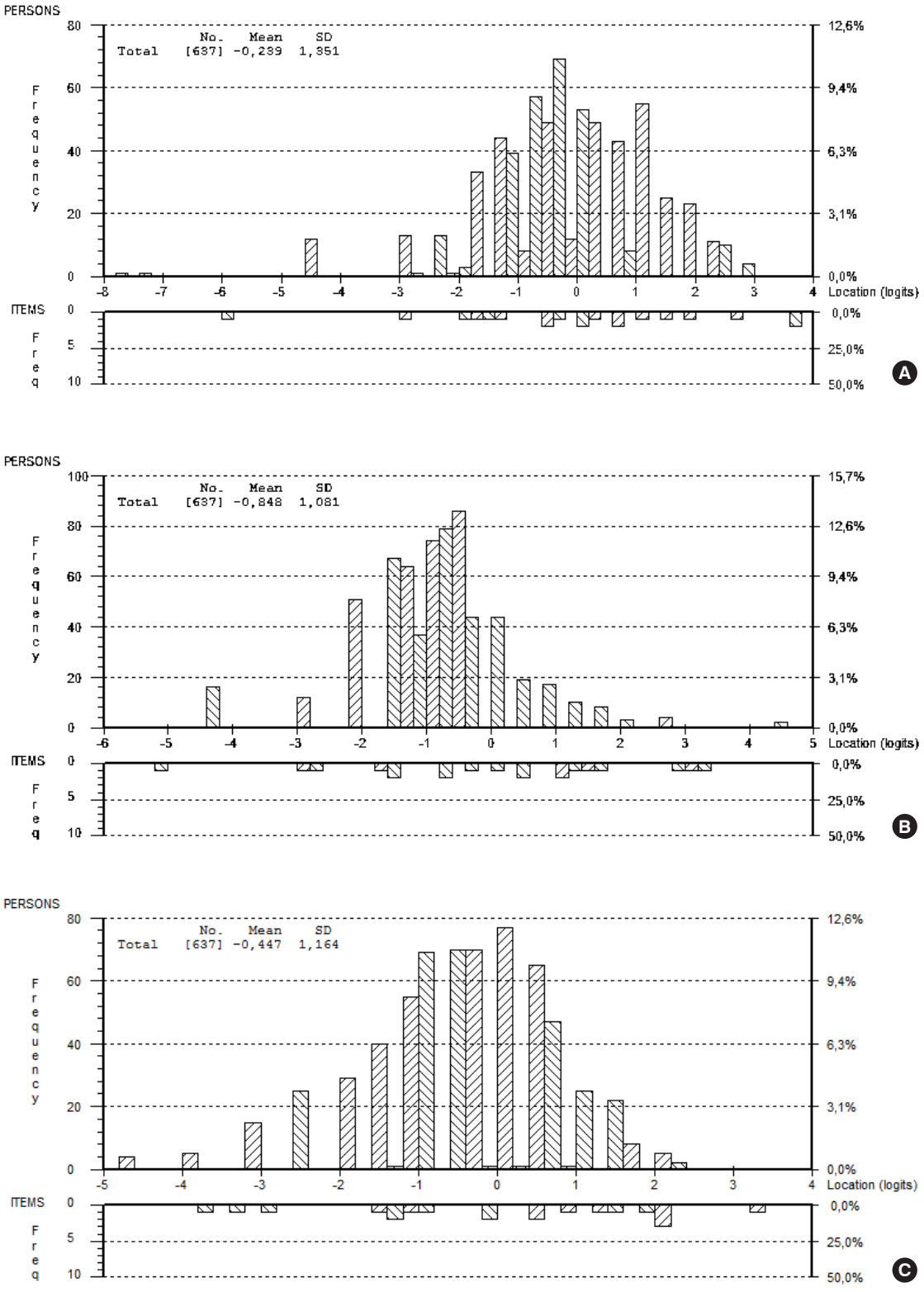


Supplementary material 3. Person-Item threshold distribution of the SRS-22r questionnaire total score. SRS-22r, revised Scoliosis Research Society-22.

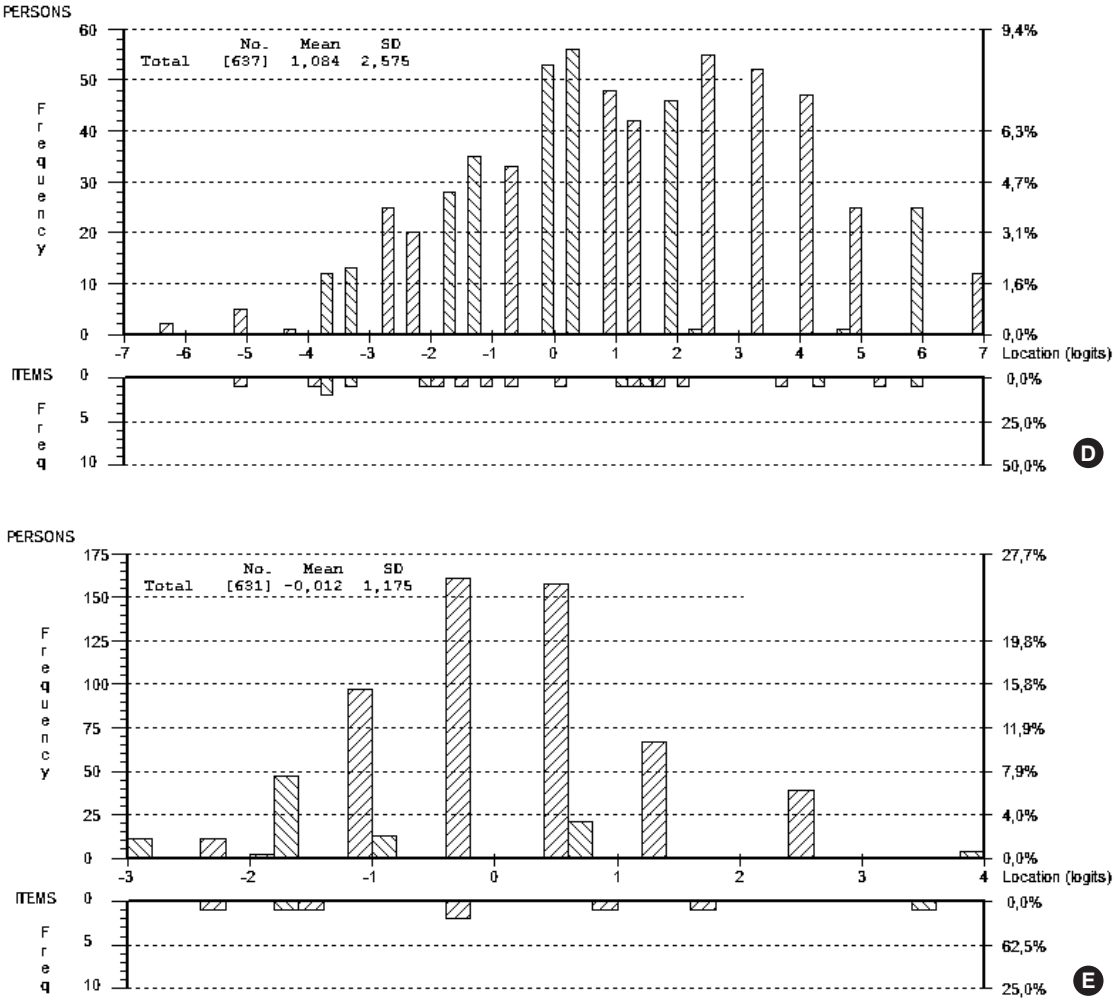


Supplementary material 4. Person-Item threshold distribution of the SRS-22r total score according to different stages of deformity. The scale provided coverage for patients locating between -4.5 and 4 logits. All patients were inside the range where the scale provided coverage. Mild or no deformity (blue), moderate (red), and marked deformity (green). SRS-22r, revised Scoliosis Research Society-22.

https://doi.org/10.14245/ns.2143354.677



Supplementary material 5. Person-Item threshold distribution of the SRS-22r domains, grouping set to interval length of 0.20 making 55 groups. Function/activity (A), pain (B), self-image (C), mental health (D), satisfaction with management (E). SRS-22r, revised Scoliosis Research Society-22. (*continued*)



Supplementary material 5. Person-Item threshold distribution of the SRS-22r domains, grouping set to interval length of 0.20 making 55 groups. Function/activity (A), pain (B), self-image (C), mental health (D), satisfaction with management (E). SRS-22r, revised Scoliosis Research Society-22. (continued)

https://doi.org/10.14245/ns.2143354.677