European Journal of Physical and Rehabilitation Medicine 2023 June;59(3):364-76
DOI: 10.23736/S1973-9087.23.07586-X

ORIGINAL ARTICLE

Cross-cultural validation of the Italian Spine Youth Quality of Life questionnaire: the ISYQOL international

Stefano NEGRINI [1, 2], Fabio ZAINA [3], Ahsen BUYUKASLAN [4, 5], Carole FORTIN [6, 7], Nikos KARAVIDAS [8], Tomasz KOTWICKI [9], Krzysztof KORBEL [10], Eric PARENT [11], Judith SANCHEZ-RAYA [12], Kathleen SHEARER [13], Hurriyet G. YILMAZ [5], Sabrina DONZELLI [3], Antonio CARONNI [14, 15] *

[1]Department of Biomedical, Surgical and Dental Sciences, University "La Statale", Milan, Italy; [2]IRCCS Istituto Ortopedico Galeazzi, Milan, Italy; [3]ISICO (Italian Scientific Spine Institute), Milan, Italy; [4]Institute of Kinesiology, University of Ljubljana, Ljubljana, Slovenia; [5]Formed Healthcare Scoliosis Brace and Treatment Center, İstanbul, Türkiye; [6]École de Réadaptation, Faculté de Médecine, Université de Montréal, Montreal, Quebec, Canada; [7]Research Center CHU Sainte-Justine, Montreal, Canada; [8]Department of Physiotherapy Schroth Scoliosis and Spine Clinic, Athens, Greece; [9]Department of Spine Disorders and Pediatric Orthopedics, Poznan University of Medical Sciences, Poznan, Poland; [10]Department of Physiotherapy, University of Medical Sciences, Poznan, Poland; [11]Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Canada; [12]Department of Physical Medicine and Rehabilitation, Hospital Universitari Vall d'Hebron, Barcelona, Spain; [13]University of Alberta, Edmonton, Canada; [14]Department of Biomedical Sciences for Health, Università degli Studi di Milano, Milan, Italy; [15]IRCCS Istituto Auxologico Italiano, Department of Neurorehabilitation Sciences, Ospedale San Luca, Milan, Italy

*Corresponding author: Antonio Caronni, Department of Neurorehabilitation Sciences, IRCCS Istituto Auxologico Italiano, via Giuseppe Mercalli 30, 20122 Milan, Italy. E-mail: a.caronni@auxologico.it

ABSTRACT

BACKGROUND: Adolescent idiopathic scoliosis and its treatments can severely impact health-related quality of life. The Italian Spine Youth Quality of Life (ISYQOL) questionnaire, initially developed in Italian and tested on Italian people, was created to measure quality of life in young persons with spine changes. ISYQOL was created using the Rasch analysis, a modern psychometric technique for questionnaires' assessment and development, which showed that the ordinal scores of the ISYQOL Italian version provide sound quality of life measures.
AIM: The current work aims to evaluate the cross-cultural equivalence of the ISYQOL questionnaire in seven different countries.
DESIGN: Cross-sectional, international, multi-centre study.
SETTING: Outpatient clinic.
POPULATION: Five hundred fifty persons with adolescent idiopathic scoliosis from English Canada, French Canada, Greece, Italy, Spain, Poland, and Türkiye.
METHODS: The ISYQOL Italian version was translated into six languages with the forward-backwards procedure. The conceptual equivalence of the items' content was verified, and any inconsistency was resolved by consensus. The Rasch analysis was used here to evaluate that ISYQOL translations retained the good measurement properties of the Italian version of the questionnaire. In addition, the Differential Item Functioning (DIF) was checked to assess the psychometric equivalence of the ISYQOL items in patients from different countries.
RESULTS: Four items of the translated ISYQOL were dropped from the questionnaire since they did not contribute to measuring due to their poor fit to the model of Rasch. Seven items were affected by DIF for nationality, a finding pointing out that these items do not work the same (*i.e.* are not equivalent) in the different countries. Thanks to the Rasch analysis, the DIF for nationality was amended, and ISYQOL International was eventually obtained.
CONCLUSIONS: ISYQOL International returns interval quality of life measures in people with adolescent idiopathic scoliosis with high cross-cultural equivalence in the tested countries.
CLINICAL REHABILITATION IMPACT: Rigorous testing showed that ISYQOL International ordinal scores return quality of life measures

cross-culturally equivalent in English and French Canada, Greece, Italy, Spain, Poland, and Türkiye. A new, psychometrically sound patient-reported outcome measure is thus available in rehabilitation medicine to measure health-related quality of life in idiopathic scoliosis.

(*Cite this article as:* Negrini S, Zaina F, Buyukaslan A, Fortin C, Karavidas N, Kotwicki T, *et al.* Cross-cultural validation of the Italian Spine Youth Quality of Life questionnaire: the ISYQOL international. Eur J Phys Rehabil Med 2023;59:364-76. DOI: 10.23736/S1973-9087.23.07586-X)

KEY WORDS: Quality of life; Psychometrics; Patient reported outcome measures; Scoliosis.

Adolescent idiopathic scoliosis is a three-dimensional change of the spine and trunk that affects up to 3% of adolescents, with females more commonly affected than males. Scoliosis and its rehabilitation treatments, *i.e.* bracing and exercise can lead to physical, psychological, and social impairments that eventually impact the quality of life.[1, 2]

Scoliosis should be diagnosed early to start appropriate monitoring and treatment.[3] Therefore, from the patient's perspective, "being involved with" scoliosis can last years; as a result, the effects of scoliosis on quality of life can also be years long. In addition, poor quality of life because of scoliosis can even persist in adulthood.[4]

Measuring quality of life is thus of the utmost importance in scoliosis, as stressed by the Physical and Rehabilitation Medicine section of the European Union of Medical Specialists (UEMS), which recommends regularly monitoring scoliosis patient-centred outcomes, including quality of life measures.[5]

Several questionnaires have been developed to evaluate the quality of life of those with spine changes.[6] The Scoliosis Research Society 22 questionnaire is the most broadly used and has been used and validated in several different languages (https://www.srs.org/professionals/online-education-and-resources/patient-outcome-questionnaires). However, it has been previously shown that the Scoliosis Research Society 22 and its variants suffer poor measurement properties.[7-10]

Recently, the Italian Spine Youth Quality of Life (ISYQOL) questionnaire has been developed.[11] ISYQOL is the first questionnaire for measuring quality of life in patients with spine changes during growth, created using the Rasch analysis. Compared to those developed with classical test theory (*i.e.* the traditional set of statistical techniques for evaluating the reliability and validity of questionnaires), a questionnaire consistent with the Rasch analysis has several strengths. Most importantly, i) its (ordinal) score can be converted into a genuine (interval) measure, and ii) this measure only reflects the quantity of the variable that the questionnaire is intended to measure (quality of life, in this case), not being distorted by confounding variables. These characteristics make a questionnaire in agreement with the Rasch analysis similar to the measuring instruments of physics (*e.g.* thermometers).

Compared to the Scoliosis Research Society 22, ISYQOL showed better psychometric properties such as better known-groups validity and can detect determinants of quality of life not demonstrated by the Scoliosis Research Society 22, such as the disease severity.[12, 13]

Several translations of ISYQOL have been published.[13-19] However, these studies only tested the translated questionnaire version in a single country. Thus, since direct comparison of the different ISYQOL translations is still missing, the actual equivalence of the questionnaire across other nations remains to be demonstrated. Moreover, in all these studies, psychometric testing was done with statistics from classical test theory, which suffers some flaws compared to Rasch analysis.

This work aims to assess the ability of ISYQOL to provide an accurate, cross-culturally equivalent measure of the quality of life in young people with idiopathic scoliosis from seven different cultures and languages (*i.e.* Canadian English, Canadian French, Greek, Italian, Polish, Spanish and Turkish). For this purpose, the Rasch analysis was used.

Materials and methods

We ran a cross-sectional study with data collected at seven specialized scoliosis clinics in six countries: Italy, Canada (French and English-speaking regions), Spain, Greece, Poland and Türkiye. All participants gave their consent to participate in the study, which was approved by the local ethical committee of each centre. Listed below are the name of the local Ethics Committees and the study's approval codes:

- Canada – English: Health Research Ethics Board, University of Alberta, Pro00073569;
- Canada – French: Sainte-Justine UHC, 2019-2275;
- Greece: National Commission for Bioethics and Technoethics, 24/10/18, Scientific Council number 02.003;
- Italy: Comitato Etico Milano Area 2, 215_2022bis;

• Poland: University of Medical Sciences, Poznan, 983/18;
• Spain: PR(AG)510/2020;
• Türkiye: Halic University Ethics Committee, 172-31.

Inclusion criteria for this study were: 1) diagnosis of adolescent idiopathic scoliosis;[1] 2) age 10 to 18 years old; 3) a frontal x-ray performed in the last six months; and 4) fluency in the tested language. Patients with idiopathic scoliosis under observation, treated with exercises or bracing were eligible. Exclusion criteria were: 1) history of spine surgery; 2) history of comorbidities in addition to scoliosis; 3) history of torso or lower extremity trauma; or 4) a positive neurological examination.

The study was undertaken in two phases. The first consisted of the forward-backwards translation of the Italian version of the ISQYOL questionnaire (*i.e.* the original one). The second consisted of administering the translated questionnaires and their Rasch analysis.

Before the expected consultation/treatment session, patients completed the ISYQOL without external help. Parents were instructed not to influence their children, and a research coordinator (a physician or a physiotherapist) was available to clarify questions and collect data.

Five hundred fifty participants were recruited in the current study, each contributing with a questionnaire. Regarding the sample size, 250 participants are usually recommended for Rasch analysis. A sample size of >500 participants returns robust estimates of the item calibrations and person measures.[20]

The ISYQOL questionnaire

ISYQOL consists of 20 questions grouped into two domains. The first 13 are completed by all patients and assess the impact of the spine condition on quality of life (spine health domain). The following seven are only relevant for brace wearers and evaluate bracing effects on quality of life (brace domain). Sixteen items investigate whether scoliosis or bracing causes a particular problem (*e.g.* item 15: "Are you worried that the brace is visible under your clothing?"). In addition, four items delineate if positive thoughts about the disease are present (e.g. Item 13: "Despite your back problem, do you live a happy life?"). Each question has three answer options (never, sometimes, often) scored from 0 to 2, and the total raw score of the whole questionnaire ranges from 0 to 40. The total score of the spine health domain ranges from 0 to 26. Sound quality of life is indicated by "never" and "often" on items investigating problems (items 1-4, 7-9, 11 and 12) and positive thoughts (items 5, 6, 10 and 13), respectively.

Quality of life is conceptualised here as a characteristic of a successful person (more is better). Therefore, the interval measure (*i.e.* the Rasch measure) obtained from the ISYQOL total score ranges from 0 to 100%, with 100% indicating the full quality of life.[11]

The cross-cultural adaptation of the ISYQOL questionnaire

The ISYQOL cross-cultural adaptation was articulated into five stages, in line with the procedure laid down by.[21] Details on the stages followed by each country are reported in Supplementary Digital Material 1, Supplementary Text File 1.

Briefly, translations were made from Italian into English, French, Greek, Polish, Spanish and Turkish according to a forward-backwards procedure. Next, a local committee reviewed, in the case, developed and eventually approved the prefinal version of the questionnaire, which was field-tested in a limited sample of respondents (10 patients) to check for difficulties and to assess understandability.[22, 23]

Regarding the forward translation, human translators were used for translating ISYQOL into French, Greek, Polish, Spanish and Turkish. Instead, the forward translation into English was supported by a software translator. On the contrary, all back translations were done by human translators. In addition, ISYQOL developers checked the back-translated versions of the questionnaire.

The seven languages versions of the final version of the ISYQOL questionnaire are provided in Supplementary Digital Material 2, Supplementary Text File 2.

The Rasch analysis of the ISYQOL questionnaire

The Rasch analysis run here (partial credit model) followed the same procedure used in previous works,[8, 11, 24, 25] and it is detailed in full in Supplementary Digital Material 3, Supplementary Text File 3.

The Rasch analysis is an iterative process consisting of different steps, assessing separate questionnaire's psychometric features. Briefly, the following ones were evaluated:

1. categories' order;
2. items' fit to the model;
3. dimensionality;
4. differential item functioning;
5. persons' reliability;
6. items' map.

As is common in Rasch analysis, if one of these param-

eters was unsatisfactory, the procedure was stopped, a solution was sought, and a new analysis ran.

Infit and outfit mean square and z-standardised statistics were calculated for each item to evaluate if each of them fitted the Rasch model. Mean squares within the 0.6 - 1.4 range suggest that data departure from the model is reasonable (*e.g.* not too large),[26] and z-standardised statistics within -1.96 and 1.96 indicate that this departure is not significant.

Dimensionality was tested by running a principal component analysis on the models' residuals. Unidimensionality was inferred if the variability taken into account by the first principal component (indicated by its eigenvalue) is small enough (<2). In the case multidimensionality is found, the procedure detailed by Smith[27] can be adopted to test if this distorts the persons' measures. Patients' measures returned by the items with positive loadings on the first principal component are contrasted to those from the items with negative loadings. Since the analysis provides patients' measures and the corresponding standard errors,[28] it is possible to test for every single person if the measurements obtained with the two sets of items are significantly different from each other. If the two measures are different in<5% of patients, multidimensionality is not considered an issue.[29]

The main aim of the current work was to evaluate if ISYQOL provides a measure of quality of life that is equivalent across cultures. Therefore, the differential item functioning for nationality was tested for each country against all countries combined. In other words, we tested if the questionnaire's items work differentially when administered to people from different countries. In the Rasch analysis, an item is affected by differential item functioning for a variable if its calibration is significantly different between two groups of participants and >0.5 logits.[30]

DIF was assessed as customary in WINSTEPS,[31] the software used here for running the Rasch analysis. First, for each of the DIF variables (*e.g.* gender), items are calibrated in the two groups of participants (*e.g.* boys and girls), and items' calibrations and calibrations standard errors are obtained. Next, the difference between the item calibrations in the two groups and the joint standard error are calculated. Finally, the DIF significance is obtained from the t-statistic (equal to the item calibrations difference/joint standard error) with joint degrees of freedom computed according to Welch-Satterthwaite. For large degrees of freedom, the t-statistic approximates the unit-normal deviate (*i.e.* the z-score).

In plain words, in this DIF analysis, a *t*-test is calculated for each item to assess if the difference in the item calibrations is statistically significant in the two groups of participants (*e.g.* boys and girls). The null hypothesis of this *t*-test is "the item has the same calibration for two groups."

The top-down purification procedure was used to identify those items that were affected by genuine DIF. According to this procedure, DIF items are deleted one at a time, starting with the items with the largest DIF.

More specifically, the items from the primary analysis are inspected for DIF. If one or more items are affected by DIF, the item with the largest DIF is identified, removed from the questionnaire, and a new analysis is run. Again, items are inspected for DIF. Thereafter, the item with the largest DIF is removed, and the analysis is rerun. This sequence is repeated until there is no longer any item with DIF. The items remaining after the top-down procedure are called "pure items" in the Rasch jargon.

The top-down purification procedure has been proposed since it has been shown that real DIF in one item favouring one group induces artificial DIF favoring the other group in the remaining items.[32] Artificial DIF is a "statistical artefact" caused by the method for detecting DIF.

Top-down purification has been applied in the current analysis, according to Lange.[33, 34]

Differential item functioning for culture and nationality is frequent, and thus it was expected for ISYQOL. In alignment with the main aim of the current work, we decided to correct any differential item functioning for nations by applying the "split items" procedure.[35] Different calibrations have been obtained for each item affected by differential item functioning, one for each group of participants. For example, if an item showed differential item functioning for nationality, with (say) Greece different from the whole sample, this item was calibrated separately for Greece and the remaining nations. In this way, the detrimental effect of the differential item functioning on measures is cancelled, and the quality of life measures in the first group of patients (Greek participants, in the current example) is safely comparable with those of the second group (the remaining nations combined) in future applications of the questionnaire.

In addition to nationality, differential item functioning was also tested for age (≤12 *vs.* >12 years), brace (not wearing *vs.* wearing the brace), disease severity (Cobb's angle ≤30° *vs.* >30°) and sex (males *vs.* females). Regarding the brace type, patients received a rigid thoracic lumbar sacral orthosis with different hours restrictions (from night to full time). Sforzesco and Sibilla braces[36] were prescribed to the Italian participants.

For practical reasons, item splitting was not used to solve any differential item functioning found for these variables. Instead, similarly to multidimensionality, it was tested if differential item functioning severely impacted measures. Also in this case, differential item functioning can be ignored if <5% of the patient's measures returned by the whole questionnaire (*i.e.* items with differential item functioning included) are significantly different from those obtained with a set of pure items (*i.e.* items free of any differential item functioning).[30, 33, 34]

Two types of DIF are commonly recognized, uniform and non-uniform, with the latter indicating an interaction between the amount of the measured variable (*e.g.* quality of life) and the DIF class (*e.g.* gender). For example, non-uniform DIF is present if an item calibration is different between boys and girls with low quality of life, but it is not between boys and girls with high quality of life. On the contrary, if uniform DIF is found, there is an item calibration difference between boys and girls, irrespectively of their quality-of-life level.

Uniform DIF flags a more substantial malfunctioning of the item. However, the study of non-uniform DIF allows for a richer DIF analysis. For the sake of simplicity, uniform DIF was only looked for in the current work.

Statistical analysis

Rasch analysis was run in WINSTEPS 5.2.5.2.[37] R 4.2.0[38] was used for additional analyses and graphics. Because of the non-normality of residuals, a Kruskal-Wallis Rank-Sum Test was preferred to regression analyses to compare age and disease severity in the different countries. Pearson's χ^2 test compared the sex distribution and the number of patients with a brace between nationality samples. The type 1 error probability (P value) was set at 0.05 except for the *t*-tests calculated for the DIF analysis. In this case, 0.01 was chosen, as done when multiple differential item functioning is tested.[33]

Results

Patients' clinical characteristics are outlined in Table I. As expected, most participants were girls and were affected by moderate/moderate-severe scoliosis, as defined using the SOSORT criteria.[1]

Age and disease severity (Cobb degrees) were significantly different in the different nation samples (Kruskal-Wallis rank-sum test: χ^2=36.2 and 54.0, respectively; df=6, P value<0.001). However, despite being significantly different, the age difference was negligible, and the difference in disease severity was rather small (Table I). In fact, the mean age ranged from 13.8 (Poland) to 15.0 years (Italy) and the mean Cobb degrees from 25.8 (French Canada) to 37.3 (Spain).

The proportion of females was not different between nations (Pearson's χ^2: χ^2=10.1, df=6, P value=0.122). On the contrary, a difference was found in the percentage of patients wearing a brace (Pearson's χ^2 test: χ^2=75.8, df=6, P value<0.001). Italy and Poland were the samples with the largest proportions of patients with a brace (about 75%), while this proportion was the smallest for French Canada (about 25%).

Between 2017 and 2019, 550 questionnaires were collected (one questionnaire per person), 250 from Italians and 50 from each remaining country. These 550 questionnaires were used for the primary analysis. In addition, a complementary analysis was also run on 350 questionnaires, with 50 Italian questionnaires randomly extracted from the 250 and 50 questionnaires from each remaining country.

Rasch analysis of the ISYQOL questionnaire

Items 5, 6, 10 and 13 of the translated questionnaires worked poorly in the Rasch measurement framework. In particular, items 5 and 13 showed disordered categories, with category 2 indicating more quality of life than

TABLE I.—*Participants' clinical characteristics.*

	Sample size	Age years	Females %	Severity Cobb degrees	Brace %
English Canada	50	13.9 (1.8)	88.0	28.3 (13.3)	42.0
French Canada	50	14.1 (1.6)	80.0	25.8 (13.9)	24.0
Greece	50	13.9 (1.8)	88.0	26.9 (9.2) [F]	46.0
Italy	250	15.0 (2.0) [E G]	87.2	32.4 (11.9) [F]	76.4
Poland	50	13.8 (1.7) [I]	90.0	33.9 (10.6) [F G]	76.0
Spain	50	14.2 (2.1)	94.0	37.3 (9.8) [E F G]	46.0
Türkiye	50	13.9 (1.8) [I]	98.0	33.8 (7.7) [E F G]	60.0

Mean (SD) is given for age and disease severity.
%: the percentage of females and patients with a brace is given. [E]significantly different from English Canada; [F]different from French Canada; [G]different from Greece; [I]different from Italy (Kruskal-Wallis rank-Sum Test followed by the Wilcoxon Rank-Sum Test with Bonferroni correction).

TABLE II.—*ISYQOL International.*

N.	Item	Categories		
		0	1	2
1	Are you afraid that your back problem may get worse?	□ never	□ sometimes	□ often
2	Are you worried about having back pain in the future because of your back problem?	□ never	□ sometimes	□ often
3	Do you feel that having your back problem is a big deal?	□ never	□ sometimes	□ often
4	Are you worried that, despite all your efforts to treat your back, it will not get better?	□ never	□ sometimes	□ often
5	Are you suffering because of your back problem?	□ never	□ sometimes	□ often
6	Does the appearance of your back make you feel uncomfortable?	□ never	□ sometimes	□ often
7	Are you worried about your back problem?	□ never	□ sometimes	□ often
8	Does it bother you to show your physical appearance?	□ never	□ sometimes	□ often
9	Are you worried that your back problem is very visible?	□ never	□ sometimes	□ often
10	Do you have to change the way that you dress because of your brace?	□ never	□ sometimes	□ often
11	Are you worried that the brace is visible under your clothing?	□ never	□ sometimes	□ often
12	Do you feel sad that you are unable to do some of the things that you used to do before you started wearing your brace?	□ never	□ sometimes	□ often
13	Do you feel your movements are restricted while wearing your brace?	□ never	□ sometimes	□ often
14	Does wearing your brace ever make you cry?	□ never	□ sometimes	□ often
15	Do you feel excluded by others because you wear your brace?	□ never	□ sometimes	□ often
16	Is wearing your brace uncomfortable?	□ never	□ sometimes	□ often

English version of the ISYQOL International questionnaire. Items 1 – 9 make the Spine-health domain. Categories "never," "sometimes" and "often" are rated 0, 1 and 2, respectively. The raw total score (0-32) is obtained by adding item scores. The raw score is then converted to a Rasch score using nation-specific score-to-measure conversion tables.

category 1 (clearly nonsensical). In addition, item 5 did not fit the model (mean square infit=1.65, z-standardized infit=6.87; mean square outfit=3.17, z-standardised outfit=9.39). After removing items 5 and 13, all the remaining items showed ordered categories. However, item 6 did not fit the model anymore (mean square outfit=1.45, z-standardised outfit=2.17), and the fit of item 10 (mean square outfit=1.37, z-standardised outfit=5.40) was significantly poor and close to the 1.4 mean square rejection threshold. The fit of item 10 remained poor (mean square outfit=1.39, z-standardised outfit=5.69) also in a subsequent analysis, in which items 5, 6 and 13 were removed.

Items 5, 6, 10 and 13 all investigate positive thoughts about the spine. Thus, items of this kind would seem to work poorly in the translated versions of the questionnaire. To simplify as much as possible the final questionnaire, all four items investigating positive thoughts (item 10 included) were removed.

The 16 remaining items, appropriately renumbered, make up the ISYQOL International (Table II) questionnaire. All the items of the ISYQOL International had ordered categories and satisfactorily fit the model.

The principal component analysis of the model's residuals showed that the eigenvalue of the first component was 2.24, which supports the notion that ISYQOL is affected by some amount of multidimensionality. However, items with positive and those with negative loadings returned significantly different measures in only 5.86% (95% CI: 3.15-9.81%, exact binomial test) of patients. This finding suggests that, for practical purposes, this small amount of multidimensionality can be ignored.

Participants' reliability was 0.80 (Cronbach's alpha of the total questionnaire score was 0.88). Therefore, ISYQOL International can distinguish ~ 3 levels of quality of life (*e.g.* low vs medium vs high) significantly different at a single subject level.

A control analysis on a sub-sample of 350 questionnaires with 50 Italian questionnaires randomly chosen confirmed these findings (see Supplementary Digital Material 4: Supplementary Text File 4).

Differential item functioning of the ISYQOL items: cross-cultural equating of the ISYQOL International questionnaire

Seven items of the ISQYOL International (items 2, 3, 4, 5, 7, 10 and 12) were corrupted by a large and significant differential item functioning for nationality (see Supplementary Text File 4 for detailed results). Given the primary goal of the current study, for each item affected by differential item functioning for nationality, parallel calibrations were calculated.

Differential item functioning was also found for brace (items 1 and 2) and sex (items 7, 10, 14 and 16). Making alternate forms of the questionnaire that also consider the differential item functioning for brace and sex (in addition to the differential item functioning for nations) would be unpractical. Thus, we opted to test the effects of the differential item functioning found for brace and sex on measures. The measures obtained by the full questionnaire

TABLE III.—*The fit of the ISYQOL International to the partial credit model.*

N.	labels	calibration	SE	IN-MNSQ	IN-ZSTD	OUT-MNSQ	OUT-ZSTD
1	Get worse	-0.63	0.09	1.09	1.42	1.09	1.37
2	Pain CF	-1.14	0.27	1.08	0.50	1.08	0.50
2	Pain	-0.01	0.09	1.15	2.44	1.18	2.77
3	Big deal CE	0.09	0.27	0.73	-1.47	0.71	-1.25
3	Big deal CF	-0.02	0.28	1.24	1.26	1.17	0.82
3	Big deal	0.99	0.09	0.93	-1.12	0.85	-1.69
4	Efforts GR	0.86	0.28	1.12	0.61	1.42	1.46
4	Efforts	0.00	0.08	1.06	1.01	1.06	0.85
5	Suffering GR	-0.75	0.28	0.99	0.03	0.98	-0.03
5	Suffering TR	-0.72	0.27	0.96	-0.19	0.93	-0.34
5	Suffering	0.92	0.10	0.88	-1.89	0.84	-1.90
6	Uncomfortable	1.09	0.09	1.00	-0.05	0.88	-1.48
7	Worried PL	-1.97	0.29	1.13	0.75	1.21	1.04
7	Worried	-0.73	0.09	1.11	1.80	1.09	1.37
8	Physical	0.70	0.08	1.07	1.14	1.14	1.44
9	Visible	0.43	0.08	1.07	1.17	0.98	-0.29
10	Brace dress PL	-0.27	0.24	0.90	-0.39	0.77	-0.66
10	Brace dress	-1.20	0.10	1.16	2.13	1.20	1.87
11	Brace visible	-1.42	0.09	1.18	2.42	1.24	2.15
12	Brace sad TR	-1.64	0.34	0.68	-1.38	0.70	-1.19
12	Brace sad	-0.12	0.10	0.94	-0.75	0.92	-0.92
13	Brace movements	-1.54	0.10	1.16	2.23	1.20	2.36
14	Brace cry	1.17	0.11	1.14	1.66	1.10	0.74
15	Brace excluded	2.37	0.14	1.26	2.23	1.30	1.30
16	Brace uncomfortable	-2.10	0.11	1.15	2.05	1.17	1.89

Note the multiple calibrations of the items affected by DIF for nationality (*e.g.* item 3).
N: item number; SE: standard error; IN-MNSQ: infit mean square; OUT-MNSQ: outfit mean square; IN-ZSTD: infit z-standardised; OUT-ZSTD: outfit z-standardized. The item's content is abbreviated by a keyword (label). CE: Canadian English; CF: Canadian French; GR: Greece; PL: Poland; TR: Türkiye. Items' calibrations and SE are given in logit.

(items with differential item functioning for brace and sex included) and those obtained with the set of items with no differential item functioning at all for brace and sex were not significantly different in any patient. Similar to multidimensionality, the differential item functioning for brace and sex does not seriously affect measures and can be ignored for practical purposes.

None of the 16 ISYQOL items was affected by differential item functioning for disease severity and patients' age.

Table III reports the fit to the model of the final version of ISYQOL International, and Figure 1 shows its items' map. The multiple calibrations of the items with differential item functioning for nationality are also given.

The different versions of ISYQOL International (*i.e.* the forms to be administered to people of different nationalities, with and without the brace) are given in Supplementary Text File 2. The corresponding score-to-measure tables are provided in Supplementary Digital Material 5: Supplementary Text File 5 and Supplementary Digital Material 6, Supplementary Text File 6. The score to measure table of the English version of ISYQOL International is given in the main text (Table IV, V).

Discussion

We tested the cross-cultural invariance of ISYQOL, a questionnaire developed with Rasch analysis to measure health-related quality of life in those with adolescent idiopathic scoliosis. When translated from Italian (*i.e.* the source language) to different languages and tested on patients from the corresponding countries (*i.e.* English Canada, French Canada, Greece, Poland, Spain and Türkiye), the ISYQOL showed some flaws. However, the Rasch analysis corrected these flaws, and the ISYQOL International questionnaire is the result. To our knowledge, ISYQOL International is the first questionnaire that provides an interval measure of quality of life in idiopathic scoliosis that is generalizable across different languages and countries.

The process through which ISYQOL International was developed starting from the original version of ISYQOL is in line with recommendations for cross-cultural adaptation of questionnaires, a two-stage process which consists of the questionnaire's translation, followed up with statistical (*i.e.* psychometric) testing.[21] Both of these stages were completed in the current work.

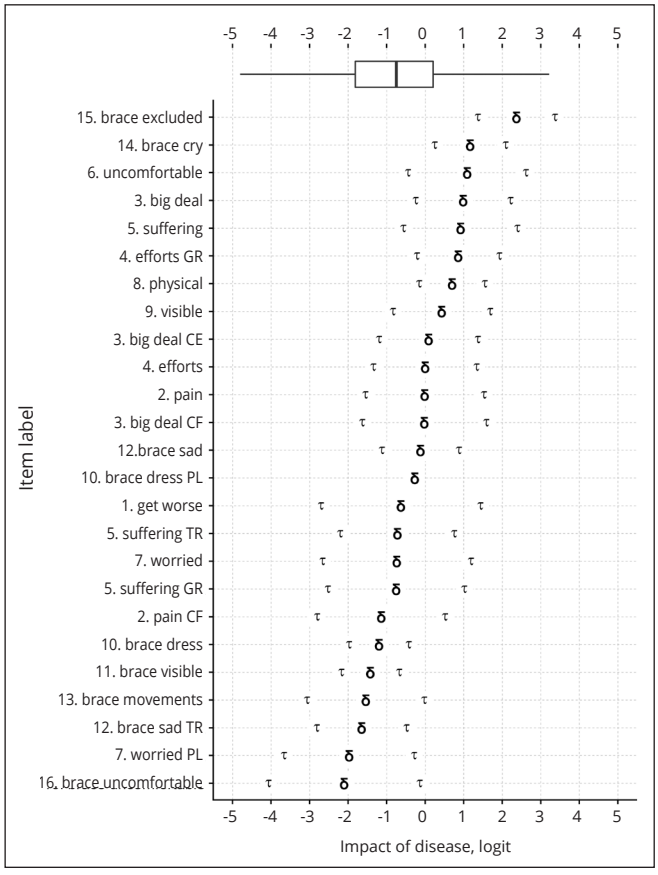

Figure 1.—Item map of ISYQOL International.
The boxplot at the top summarises the patients' measures, while the graph shows the items' calibration.
CE: Canadian English; CF: Canadian French; GR: Greece; PL: Poland; TR: Türkiye; δ: item's mean measure. τ: Andrich's thresholds (the leftmost τ marks the boundary between category 0 and 1, while the rightmost one that between 1 and 2); τ are not shown for "item 10 PL" since this is the only item with disordered thresholds. Notably, all items ("item 10 PL" included) had ordered categories. Patients' measures and items' calibrations are referred to an interval scale with 1 logit as the measurement unit (*i.e.* horizontal axis) and centred on 0, *i.e.* the items' mean calibration. Moving from negative (*e.g.*, -5) to positive (*e.g.*, +5) logit, the impact of the disease and its treatments in the patient's life gets more and more disturbing (and thus the patient's quality of life gets worse and worse).

TABLE IV.—*Conversion table of the raw scores of the ISYQOL International (English version) to the Rasch interval measure.*

Score	Measure (%)	SE (%)
0	100.00	15.79
1	88.79	9.13
2	81.52	6.85
3	76.78	5.86
4	73.13	5.27
5	70.09	4.87
6	67.43	4.59
7	65.06	4.37
8	62.87	4.20
9	60.84	4.08
10	58.91	3.98
11	57.07	3.90
12	55.29	3.84
13	53.55	3.80
14	51.85	3.77
15	50.17	3.75
16	48.50	3.75
17	46.83	3.75
18	45.16	3.75
19	43.48	3.77
20	41.78	3.80
21	40.04	3.84
22	38.27	3.89
23	36.43	3.97
24	34.51	4.07
25	32.48	4.21
26	30.29	4.39
27	27.87	4.64
28	25.11	5.01
29	21.81	5.57
30	17.51	6.55
31	10.78	8.86
32	0.00	15.61

Score: total score of the whole questionnaire. Measures: interval measures on a scale ranging from 0 to 100, the latter indicating complete quality of life. SE: standard error (*i.e.* an estimate of the uncertainty of the measure). This score-to-measure conversion table is used by patients wearing a brace, *i.e.* the patients who answered the whole questionnaire (16 items).

The final aim of the adaptation process was to reach a full equivalence between the original and the new version(s) of the questionnaire.[21] Full equivalence means that alternate versions measure the same variable in the same way (*i.e.* they have the same psychometric properties).

Based on that, the translation should ensure the conceptual equivalence between the different questionnaire versions and their semantic equivalence (*i.e.* equivalence in the meaning of the words).[21, 39] The conceptual equivalence of the translated items of ISYQOL was checked here by the ISYQOL developers, who discussed with the local research team any inconsistencies with the source questionnaire.

The differential item functioning analysis is a powerful tool for assessing psychometric equivalence in the cross-cultural adaptation of questionnaires.[40, 41] This assessment was run here in the framework of the Rasch analysis. Still, it should be pointed out that the differential item functioning can also be evaluated without resorting to the model of Rasch.[42]

Seven out of 16 items of ISYQOL international were affected by differential item functioning for nationality. However, differential item functioning is quite common when several cultures are compared,[35] and the testing of the differential item functioning was substantial herein.

TABLE V.—*Score-to-measure conversion table of the Spine health domain of the ISYQOL International (English version).*

Score	Measure (%)	SE (%)
0	100.00	18.49
1	86.73	10.82
2	77.83	8.30
3	71.74	7.27
4	66.82	6.69
5	62.52	6.33
6	58.61	6.09
7	54.93	5.93
8	51.42	5.83
9	48.01	5.75
10	44.67	5.72
11	41.34	5.73
12	37.97	5.80
13	34.46	5.97
14	30.66	6.27
15	26.33	6.82
16	20.93	7.87
17	12.75	10.47
18	0.00	18.30

Score: total score of the Spine-health domain of the ISYQOL questionnaire (Items 1 to 9). Remaining abbreviations as in Table IV. This score-to-measure conversion table is for patients not wearing a brace, who only completed the first nine items of the questionnaire (*i.e.* the Spine-health domain).

Indeed, differential item functioning should always be suspected when a questionnaire is translated into different languages and administered to people from other countries and cultures.

The differential item functioning represents a violation of unidimensionality, which is an assumption of measurement and an assumption of the Rasch analysis.[43] The score of an item affected by differential item functioning depends not only on the variable the questionnaire aims to measure (*e.g.* quality of life) but also on a secondary variable (*e.g.* nationality). Of course, quality of life may differ for people from two different countries, and thus the questionnaires' scores can be different. The differential item functioning points out that the questionnaires scores are different in these two persons independently of a difference in their quality of life level. If this is the case, the persons' scores (and thus their measures) are different, but the real value of the variable is not.

If the differential item functioning is present, its impact on measurement may be minor from a practical point of view. The artefact caused by the differential item functioning can be quantified and if it is shown to be minor, the differential item functioning can be eventually ignored,[30] which is ultimately how we were able to navigate the observed differential item functioning for sex and brace. However, the ultimate goal of the current work is to provide a questionnaire for measuring quality of life in adolescent idiopathic scoliosis that can be appropriately and safely used across different cultures. For this reason, we preferred to use the split items procedure[44] to fully compensate for any differential item functioning for nationality, irrespective of its size.

By implementing the split items procedure, the Rasch analysis offers an advantageous way to amend the lack of cross-cultural invariance pointed out by the differential item functioning. Therefore, assessing the differential item functioning in the Rasch analysis framework identifies the problem with the cross-cultural validity of a questionnaire and offers a solution to this problem.

Regarding the alternate forms of the score-to-measure tables, it is important to stress that these are provided (for the different countries) for the spine health domain and the entire questionnaire. Thus, ISYQOL International shares a real strength with ISYQOL original: they both allow a direct comparison of quality of life in people with and without the brace.[11, 12]

The conversion of the questionnaire total scores into interval measures deserves some additional comments. Consider a Canadian patient without the brace who speaks English. This patient scores 2 on the Spine-health domain of ISYQOL International. From Table V, her measure of quality of life is 77.83% (±8.30%) of the maximum quality of life detected by the questionnaire. A brace is prescribed for this girl, and her quality of life is measured some months later. Now she fills out the complete questionnaire, and her total score is 8. From Table IV, the corresponding quality of life measure is 62.87% (±4.20%). This example clearly shows that questionnaire scores are misleading. After the brace prescription, the patient's total score passes from 2 to 8, *i.e.* the problems decreasing the quality of life become four times larger. However, the quality of life measure only drops from 77.83% to 62.87%.

In addition, it can be shown that the worsening of the girl's quality of life (*i.e.* 14.96) is not significant. In fact, under the null hypothesis that the two measures are the same: the difference between the two measures is 77.83-62.87=14.96; the standard error of the difference is: $\sqrt{(8.3^2+4.2^2)}=9.30$ and the 95% confidence interval of the difference between the two measures is 14.96±1.96 * 9.30 (*i.e.* from -3.27 to 33.19), which includes 0.

From another point of view, the modification of the girl's quality of life is not significant since her Rasch change index is<1.96 (Rasch change index: 14.96 / 9.30=1.61).[28]

It is worth doing another reflection on the conversion from the ordinal score to the interval measure. ISYQOL International collects the problems caused to patients by the

disease and bracing. The more the problems, the higher the total questionnaire score, and the lower the quality of life. This relationship still holds when the total questionnaire score is converted in the logit measure. The higher (*i.e.*, the more positive) the logit measure, the lower the quality of life, as shown in Figure 1 and Supplementary Text File 6. However, we overturned the relationship between measures and quality of life, when these are expressed on the 0-100 scale. In this case, the higher the measure, the better the quality of life. This choice was dictated by the fact that quality of life is a positive construct (the higher, the better), and we feel more meaningful that a scale aimed at measuring quality of life also reflects this. This should also be more acceptable in a clinical context. Full quality of life is indicated by 100 (out of 100). As the disease and bracing cause more and more problems, the measure of quality of life is reduced.

In the current analysis, four items of the translated version of ISYQOL original poorly fitted the model of Rasch, a finding suggesting that their ability actually to contribute to measurement should be questioned. We feel that this malfunctioning is most likely attributable to linguistic rather than cultural reasons.

Items with a poor fit were related to positive thoughts about scoliosis. Items of this type were included in the questionnaire since it was previously shown that patients could also find positive aspects of their disease and back condition.[11] We believe that these items are too convoluted when translated from Italian into different languages.[21] Their phrasing is "too Italian," and they do not work anymore in other languages. In addition, the lack of idiomatic equivalence[21] cannot be completely ruled out. For example, item 10 of ISYQOL original contains a colloquialism ("non è una tragedia"; in English: "it's no big deal"). In the forward-backwards procedure, equivalent expressions in the target languages have been looked for. Still, the chance that this item did not retain a similar meaning in the different translations remains.

Because of the poor fit to the model and based on this reasoning, we ultimately chose to remove them. In so doing, the process required to achieve the questionnaire's total score summation procedure was simplified, given that the items investigating positive thoughts needed to be recoded before being added to the remaining items.[11]

Regarding sample size, it should be noted that there was a clear predominance of Italian questionnaires, which was done for a purposeful methodological reason. The ISYQOL was developed to measure quality of life in young people with spine changes. The questionnaire assumes that if disease and its treatments cause problems to patients, their quality of life is decreased. For this reason, our group chose to collaborate directly with Italian adolescents with scoliosis and kyphosis to identify from their perspective the problems related to their disease through the use of interviews.[11] Next, a content analysis was run on the participants' transcripts, and a set of potential questionnaire items emerged. Finally, expert clinicians were asked to choose the items best suited to measure the quality of life in idiopathic scoliosis. It must be stressed that this whole process, which was run in Italian and with Italian patients, gives high validity to the ISYQOL as a quality of life measurement.[45]

Biasing the sample size toward Italy ensures that the framework in which the measurement of quality of life in spinal deformities was initially developed is maintained throughout the assessment of the cross-cultural equivalence of the questionnaire (*i.e.* the DIF analysis). In a sense, the comparison is between the translated and the Italian versions, with the six translations conforming to the source questionnaire. In this "paired" comparison, it is noteworthy that the number of Italian questionnaires recorded is about as numerous as those completed by the remaining countries put together.

In addition, to check that the difference in the sample size did not bias the study's main findings, a complementary analysis was also run on 350 questionnaires, with a sample of 50 questionnaires randomly extracted from the 250 Italian ones and 50 questionnaires for each of the remaining countries. The complete analysis and complementary returned the same results (Supplementary Text File 4).

It is almost superfluous to stress that assessing the cross-cultural invariance of questionnaires is of paramount importance. In that regard, it suffices to recall that any pooling of data in international studies implicitly assumes that the different translations of the questionnaire are equivalent in the different cultures.[35] Similarly, cross-cultural equivalence is also presumed when studies' results are pooled at a meta-analytic level.[46]

The current one is not the first study in which a cross-cultural evaluation of ISYQOL has been provided. English,[13] French[19] and Polish[14] versions of ISYQOL are already available, and ISYQOL has also been translated into languages not tested here, such as Chinese,[15] Korean,[18] Persian[16] and Arabic.[17] According to these works, ISYQOL works well in its current form, a conclusion in contradiction with the current analysis findings. The explanation for this can be twofold. First, to our knowledge, this is the first study in which several translations were simultaneously tested and compared with the original version of

the questionnaire. Second, this is the first study that used the Rasch analysis, whereas statistics from the classical test theory were adopted in previous reports. The different results related to the cross-cultural validity of ISYQOL are not surprising if the methodological differences between item response theory (Rasch analysis included) and classical test theory are taken into account.[47]

Regarding the ISYQOL translations and dissemination, it is also noteworthy that a recent literature review endorsed ISYQOL as the appropriate patient-reported outcome measure for measuring the quality of life of young people with idiopathic scoliosis or kyphosis who are under medical observation or brace treatment.[6]

Limitations of the study

We are aware of some of the limitations of our study. Firstly, even if we tested the cross-cultural equivalence of ISYQOL in seven different countries, the cross-cultural equivalence of the questionnaire could still be assessed in greater detail. For example, the English version of ISYQOL International could be tested on people from the United Kingdom or the United States and French on people from France. In these cases, the linguistic equivalence of ISYQOL International can be reasonably assumed, but the cultural one remains to be evaluated.

Regarding the ISYQOL cross-cultural adaptation process, one could consider it suboptimal that a software supported the forward translation from Italian into English. However, it has been experimentally shown that while fully automated translations from machine translators are weak, semi-automated translations (*i.e.*, later reviewed and refined by humans, as done here) can be considered acceptable. For example, when machine translation is followed by human editing, the translation of health documents is comparable to that obtained with human translation only.[48] Intriguingly, machine translations could even have some strengths over human translations. For example, they are said to ensure a translation that reflects a wide range of translators instead of the experience of two only, as is usually done.[49]

Still, on the subject of the forward translation of a questionnaire, it should also be stressed that this is only the first step of the questionnaire's cross-cultural adaptation.[21] The forward translation cannot alone ensure or prevent equivalence between the translated and the source version. Instead, researchers decide equivalence by consensus. To this aim, they can (and should) refine the translation as needed.

The psychometrics functioning of ISYQOL International also remains to be tested in different patient populations, common in the spine clinic.

For example, further questionnaire testing could be done in persons waiting for spinal surgery and with a history of spinal surgery. Since ISYQOL International does not directly ask about surgery-related problems, rightly one may ask if this questionnaire is suitable for measuring the quality of life in this patient group. However, in the Rasch analysis framework, items with very different content may mark the same level of quality of life. Therefore, the quality of life quantity assessed by a hypothetical item investigating a surgery-related problem could be already probed by one of the ISYQOL International items. Of course, despite this theoretical argument, the actual functioning of ISYQOL International should be field-tested with dedicated data collection.

Similarly, field testing of ISYQOL International in patients with other spine conditions than idiopathic scoliosis, such as Scheuermann juvenile kyphosis in the first place, remains to be done. In this regard, it is worth noting that the ISYQOL Italian version[11] has been developed for scoliosis and kyphosis persons. ISYQOL Italy showed measurement invariance in these two spine conditions indicating that it provides measures suitable for comparing the quality of life in the two primary diagnostic groups of idiopathic spine diseases.[12] This finding bodes well also for ISYQOL International.

In any case, as done before, the DIF analysis would be the right tool for studying the stability of ISYQOL International in scoliosis and kyphosis as well as in persons who had and those who had not had surgery and in other contrasts of clinical interest.

Concerning the disease progression and its treatments, ISYQOL International's responsiveness and its measurement properties' invariance over time should also be evaluated.[25, 28, 50]

Conclusions

ISYQOL International is a 16 items questionnaire (each scored in three categories) to measure health-related quality of life in young people with idiopathic scoliosis. ISYQOL International was developed with the Rasch analysis starting from the ISYQOL Italian version. Rigorous testing showed that ISYQOL International ordinal scores return sound quality of life measures. In addition to being psychometrically sound, these measures are also cross-culturally equivalent in English Canada, French Canada, Greece, Italy, Spain, Poland, and Türkiye.

References

1. Negrini S, Donzelli S, Aulisa AG, Czaprowski D, Schreiber S, de Mauroy JC, *et al.* 2016 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis Spinal Disord 2018;13:3.

2. Freidel K, Petermann F, Reichel D, Steiner A, Warschburger P, Weiss HR. Quality of life in women with idiopathic scoliosis. Spine 2002;27:E87–91.

3. Scaturro D, de Sire A, Terrana P, Costantino C, Lauricella L, Sannasardo CE, *et al.* Adolescent idiopathic scoliosis screening: could a school-based assessment protocol be useful for an early diagnosis? J Back Musculoskeletal Rehabil 2021;34:301–6.

4. Schroeder JE, Michaeli T, Luria M, Itshayek E, Kaplan L. Long-Term Effects on Sexual Function in Women Treated With Scoliosis Correction for Adolescent Idiopathic Scoliosis. Global Spine J 2022;21925682221079263.

5. Negrini S, Dincer F, Kiekens C, Kruger L, Varela-Donoso E, Christodoulou N. Evidence based position paper on physical and rehabilitation medicine (PRM) practice for people with spinal deformities during growth. The European PRM position (UEMS PRM Section). Eur J Phys Rehabil Med 2017;53:125–31.

6. Babaee T, Moradi V, Shariat A, Anastasio AT, Khani A, Bagheri M, *et al.* Disease-Specific Outcome Measures Evaluating the Health-Related Quality of Life of Children and Adolescents with Idiopathic Scoliosis and Scheuermann's Kyphosis: A Literature Review. Spine Surg Relat Res 2022;6:197–223.

7. Rothenfluh DA, Neubauer G, Klasen J, Min K. Analysis of internal construct validity of the SRS-24 questionnaire. Eur Spine J 2012;21:1590–5.

8. 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.

9. Aksekili MA, Demir P, Iyigun A, Akcaalan S, Korkmazer S, Tecimel O, *et al.* Turkish Validity and Reliability Study of Scoliosis Research Society-30 Questionnaire in Adolescent Idiopathic Scoliosis Patients. Spine 2021;46:E1058–64.

10. Parent EC, Dang R, Hill D, Mahood J, Moreau M, Raso J, *et al.* Score distribution of the scoliosis research society-22 questionnaire in subgroups of patients of all ages with idiopathic scoliosis. Spine 2010;35:568–77.

11. Caronni A, Sciumè L, Donzelli S, Zaina F, Negrini S. ISYQOL: a Rasch-consistent questionnaire for measuring health-related quality of life in adolescents with spinal deformities. Spine J 2017;17:1364–72.

12. Caronni A, Donzelli S, Zaina F, Negrini S. The Italian Spine Youth Quality of Life questionnaire measures health-related quality of life of adolescents with spinal deformities better than the reference standard, the Scoliosis Research Society 22 questionnaire. Clin Rehabil 2019;33:1404–15.

13. Alanazi MH, Parent EC, Bettany-Saltikov J, Hill D, Southon S. Convergent validity, ceiling, and floor effects of the English-ISYQOL against established quality of life questionnaires (SRS-22r and SAQ) and curve angles in adolescents with idiopathic scoliosis. Stud Health Technol Inform 2021;280:225–30.

14. Kinel E, Korbel K, Janusz P, Kozinoga M, Czaprowski D, Kotwicki T. Polish Adaptation of the Italian Spine Youth Quality of Life Questionnaire. J Clin Med 2021;10:2081.

15. Liu S, Liang J, Xu N, Mai S, Wang Q, Zeng L, *et al.* Reliability and validity of simplified Chinese version of the Italian spine youth quality of life questionnaire in adolescents with idiopathic scoliosis. BMC Musculoskelet Disord 2021;22:568.

16. Babaee T, Esfandiari E, Rouhani N, Nakhaee M, Saeedi M, Hedayati Z, *et al.* The Italian Spine Youth Quality of Life questionnaire: reliability and validity of the Persian version in adolescents with spinal deformities. Spine Deform 2022;10:775–82.

17. Fallatah SM, Emam S, Al-Ghamdi G, Almatrafi F. Cross-cultural adaptation and validation of the Italian Spine Youth Quality of Life (ISYQOL) questionnaire's Arabic version. Medicine (Baltimore) 2021;100:e28063.

18. Park SH, Goh TS, Son SM, Kim DS, Lee JS. Validation of the Italian spine youth quality of life (ISYQOL) in Korean population. J Clin Neurosci 2021;92:165–8.

19. Forest C, Parent E, Chémaly O, Barchi S, Donzelli S, Negrini S, *et al.* Cross-cultural French-Canadian adaptation and psychometric assessment of the Italian Spine Youth Quality of Life (ISYQOL) questionnaire. Spine J 2022;22:1893–902.

20. Linacre JM. Sample size and item calibration (or person measure) stability. Rasch Meas Trans 1994;328.

21. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000;25:3186–91.

22. Hambleton RK, Merenda PF, Spielberger CD. Adapting educational and psychological tests for cross-cultural assessment. Mahwah, NJ: Lawrence Erlbaum; 2004.

23. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, *et al.* Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002;11:193–205.

24. Negrini S, Donzelli S, Di Felice F, Zaina F, Caronni A. Construct validity of the Trunk Aesthetic Clinical Evaluation (TRACE) in young people with idiopathic scoliosis. Ann Phys Rehabil Med 2020;63:216–21.

25. Caronni A, Picardi M, Redaelli V, Antoniotti P, Pintavalle G, Aristidou E, *et al.* The Falls Efficacy Scale International is a valid measure to assess the concern about falling and its changes induced by treatments. Clin Rehabil 2022;36:558–70.

26. Wright B. Reasonable mean-square fit values. Rasch Meas Trans 1994;8:370.

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

28. Caronni A, Picardi M, Gilardone G, Corbo M. The McNemar Change Index worked better than the Minimal Detectable Change in demonstrating the change at a single subject level. J Clin Epidemiol 2021;131:79–88.

29. Tennant A, Pallant JF. Unidimensionality matters!(A tale of two Smiths?). Rasch Meas Trans 2006;20:1048–51.

30. Tennant A, Pallant J. DIF matters: A practical approach to test if differential item functioning makes a difference. Rasch Meas Trans 2007;20:1082–4.

31. Linacre JM. DIF - DPF - bias - interactions concepts. Winsteps Help for Rasch Analysis, [Internet]. Available from: https://www.winsteps.com/winman/difconcepts.htm [cited 2023, Apr 26].

32. Andrich D, Hagquist C. Real and Artificial Differential Item Functioning in Polytomous Items. Educ Psychol Meas 2015;75:185–207.

33. Lange R, Irwin HJ, Houran J. Top-down purification of Tobacyk's Revised Paranormal Belief Scale. Pers Individ Dif 2000;29:131–56.

34. Lange R, Thalbourne MA, Houran J. Depressive response sets due to gender and culture-based differential item functioning. Pers Individ Dif 2002;33:937–54.

35. Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, *et al.* Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004;42(Suppl):I37–48.

36. Negrini S, Marchini G, Tessadri F. Brace technology thematic series - The Sforzesco and Sibilla braces, and the SPoRT (Symmetric, Patient oriented, Rigid, Three-dimensional, active) concept. Scoliosis 2011;6:8.

37. Linacre JM, Wright BD. Winsteps [Internet]. Available from: http://www winsteps com/index htm [cited 2013, Jun 27].

38. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.

39. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417–32.

40. Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Power M, *et al.*; KIDSCREEN Group. The KIDSCREEN-52 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. Value Health 2008;11:645–58.

41. Catz A, Itzkovich M, Tesio L, Biering-Sorensen F, Weeks C, Laramee MT, *et al.* A multicenter international study on the Spinal Cord Independence Measure, version III: rasch psychometric validation. Spinal Cord 2007;45:275–91.

42. Shealy R, Stout W. A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. Psychometrika 1993;58:159–94.

43. Bond TG, Fox CM. Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah, NJ: Lawrence Erlbaum; 2013.

44. Roorda LD, Jones CA, Waltz M, Lankhorst GJ, Bouter LM, van der Eijken JW, *et al.* Satisfactory cross cultural equivalence of the Dutch WOMAC in patients with hip osteoarthritis waiting for arthroplasty. Ann Rheum Dis 2004;63:36–42.

45. Sireci SG. The Construct of Content Validity. Soc Indic Res 1998;45:83–117.

46. Li X, Shen J, Liang J, Zhou X, Yang Y, Wang D, *et al.* Effect of core-based exercise in people with scoliosis: A systematic review and meta-analysis. Clin Rehabil 2021;35:669–80.

47. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009;13:iii, ix–x, 1–177.

48. Turner AM, Bergman M, Brownstein M, Cole K, Kirchhoff K. A comparison of human and machine translation of health promotion materials for public health practice: time, costs, and quality. J Public Health Manag Pract 2014;20:523–9.

49. Koehn P. Statistical machine translation. Cambridge: Cambridge University Press; 2009.

50. Caronni A, Sciumè L. Is my patient actually getting better? Application of the McNemar test for demonstrating the change at a single subject level. Disabil Rehabil 2017;39:1341–7.

Conflicts of interest
The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

Funding
For the ISYQOL French version, Claudie Forest received a grant awarded by the Programme d'Excellence en Médecine pour l'Initiation En Recherche (PREMIER) from the School of Rehabilitation at University of Montreal and Carole Fortin was supported by a Junior 1 salary award from Fonds de Recherche du Québec - Santé (FRQS). Antonio Caronni was supported by the IRCCS Istituto Auxologico Italiano – Milano, Italia, within the RESET research project (Ricerca Corrente 2020, Italian Ministry of Health).

Authors' contributions
Stefano Negrini conceived the study; Stefano Negrini, Fabio Zaina, Eric Parent and Antonio Caronni contributed to the study's design; Fabio Zaina, Ahsen Buyukaslan, Carole Fortin, Nikos Karavidas, Tomasz Kotwicki, Krzysztof Korbel, Eric Parent, Judith Sanchez-Raya, Kathleen Shearer, Hurriyet G. Yilmaz and Sabrina Donzelli, *i.e.* the local researchers, coordinated the local data collection, contributed to collecting data and arranged the local datasets; Fabio Zaina and Sabrina Donzelli arranged the local datasets in a whole dataset; Antonio Caronni analyzed the whole dataset, prepared the figures and wrote the first draft of the manuscript; Antonio Caronni and Stefano Negrini discussed the first manuscript draft. After that, all authors contributed to revising a subsequent draft version. Antonio Caronni arranged the final version of the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgments
We are profoundly indebted with: Andrea Lin (who received a Women and Children Health Research Institute Summer studentship, for working on part of this project), Kyle Stampe, Eric Huang, Jim Mahood, Marc Moreau and Sarah Southon Hryniuk (for assisting with recruitment in Edmonton, Canada); Claudie Forest (for contributing to the development of the French version of the questionnaire), Resi Contardo (for the Italian to French translation) and the Orthopedics Clinical Research Unit team at Sainte-Justine University Hospital Center (who contributed to the recruitment of the participants in Montreal, Canada); Papastefanou Sotirios and Sekouris Nikolaos (for blindly measuring the scoliosis severity on the X-rays); Edyta Kinel, Mateusz Kozinoga, Dariusz Czaprowski, Piotr Janusz, Katarzyna Politarczyk (for contributing to the development of the Polish version of the questionnaire and assisting with recruitment); Elisabetta D'Agata (for the Italian to Spanish translation of the questionnaire), Clara Figueras (for assisting with recruitment), Joan Bagó and Manuel Rigo (for contributing to the development of the final version of the Spanish questionnaire); Tugba Kuru Çolak (for her support), Nurperi Gazioğlu and Sintiya Levi (for the Italian to Turkish translation of the questionnaire).

History
Article first published online: May 17, 2023. – Manuscript accepted: April 19, 2023 - Manuscript revised: November 14, 2022. - Manuscript received: May 9, 2022.

Supplementary data
For supplementary materials, please see the HTML version of this article at www.minervamedica.it