Accepted author's manuscript. Published in final edited form as:
Spine Deformity 2022; 10(5): 1055-1062. Publisher DOI: 10.1007/s43390-022-00509-5

Shorter and Sweeter: the 16-Item Version of the SRS-Questionnaire Shows Better Structural Validity Than the 20-Item Version in Young Patients With Spinal Deformity

F. Mannion[1] · A. Elfering[2] · T. F. Fekete[3] · I. J. Harding[4] · M. Monticone[5] · P. Obid[6] · T. Niemeyer[7] · U. Liljenqvist[8] ·A. Boss[9] · L. Zimmermann[1] · A. Vila-Casademunt[10] · F. J. Sánchez Perez-Grueso[11] · J. Pizones[11] · F. Pellisé[12] ·S. Richner-Wunderlin[1] · F. S. Kleinstück[3] · I. Obeid[13] · L. Boissiere[13] · A. Alanay[14] · J. Bagó[10]

[1] Department Teaching, Research and Development, Spine Subdivision, Schulthess Klinik, Lengghalde 2, 8008 Zurich, Switzerland
[2] Institute of Psychology, University of Bern, Bern, Switzerland
[3] Spine Centre, Schulthess Klinik, Zurich, Switzerland
[4] North Bristol NHS Trust, Bristol, UK
[5] Department Medical Sciences and Public Health, University of Cagliari, Cagliari, Italy
[6] Dept. of Orthopaedics and Orthopaedic Surgery, Greifswald University Hospital, Ferdinand-Sauerbruch-Strasse, 17475 Greifswald, Germany
[7] Spine and Scoliosis Center, Asklepios Paulinen Klinik Wiesbaden, Wiesbaden, Germany
[8] Department Spine Surgery, St Franziskus Hospital, Münster, Germany
[9] Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
[10] Spine Research Unit, Vall d'Hebron Institute of Research, Barcelona, Spain
[11] Spine Unit, Department of Orthopedic Surgery, University Hospital La Paz, Madrid, Spain
[12] Spine Unit, Hospital Vall d'Hebron, Barcelona, Spain
[13] Spine Surgery Unit, Pellegrin University Hospital, Bordeaux, France
[14] Dept. Orthopedics and Traumatology, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

Corresponding author: A. F. Mannion anne.mannion@yahoo.com

Purpose In patients with adult spinal deformity, it was previously shown that 16 of the non-management items of the SRS-instrument showed a better fit to the theoretical four-factor model (pain, function, self-image, mental health) than did all 20 items. Whether the same phenomenon is observed in data from younger (<20y) patients, for whom the questionnaire was originally designed, is not currently known.

Methods Confirmatory factor analysis was used to evaluate the factor structure of the 20 non-management items of the SRS-instrument completed by 3618 young patients with spinal deformity (75.5% female; mean age,15.0±2.0y) and of its equivalence across language versions (2713 English-speaking, 270 Spanish, 264 German, 223 Italian, and 148 French). The root mean square error of approximation (RMSEA) and comparative fit index (CFI) indicated model fit.

Results Compared with the 20-item version, the 16-item solution significantly increased the fit ($p<0.001$) across all language versions, to achieve good model fit (CFI=0.96, RMSEA=0.06). For both 16-item and 20-item models, equivalence across languages was not reached, with some items showing weaker item-loading for some languages, in particular German and French.

Conclusion Also in patients with adolescent idiopathic scoliosis, the shorter 16-item version showed a better fit to the intended 4-factor structure of the SRS-instrument. The wording of some of the items, and/or their equivalence across language versions, may need to be addressed. Questionnaire completion can be a burden for patients; if a shorter, more structurally valid version is available, its use should be encouraged.

Key words: SRS-22, Patient Reported Outcome, Factor analysis, language versions, AIS

Introduction

The SRS-22R was developed to assess health-related quality of life (HRQL) in patients with adolescent idiopathic scoliosis and is one of the most widely used patient-rated outcome measures (PROMs) in clinical studies investigating spinal deformity. It is a multidimensional instrument covering four non-management domains, including pain, function, self-image, and mental health, with 5 items per domain, along with 2 items to assess the patient's satisfaction with management of their condition. The domains selected for inclusion intend to capture the wide-ranging impact of scoliosis on various aspects of the patient's life. Weaknesses identified in the factor structure of the instrument [1-3], led to the proposal of a shorter, unidimensional, and linearly-scaled 7-item version (SRS-7) of the instrument [4]. However, the authors conceded that, while this was useful for assessing global changes in patient outcomes over time, a multidimensional instrument was likely better for assessing changes associated with individual aspects of the disease [4]. The SRS-7 does not appear to have entered into common use. A systematic review of the psychometric properties of 17 published translations ("official versions") of the SRS-22 instrument, carried out more than a decade ago, concluded that further attention should be given to the construct validity of the SRS-22 in terms of its cross-cultural validity (i.e., the extent to which the performance of an item in an adapted instrument adequately reflects its performance in the original version) and structural validity (i.e., the extent to which the scores adequately reflect the dimensional nature of the construct being measured)[5]. In response to this, in 2018 the factor structure of different language versions of the instrument was investigated, but this was done in older patients with adult spinal deformity [6]. It was reported that 16 of the non-management items of the SRS-22 showed a significantly better fit to the theoretical

four-factor model (pain, function, self-image, mental health) than did all 20 items. The worst-fitting item per domain was recommended for exclusion (respectively: Q17, sick days; Q15, financial difficulties; Q14 personal relationships; Q3, nervous). Issues related to the fit of some of these same items had also been reported incidentally during the development of other language versions of the instrument, tested in young patients with AIS [7, 8]. This suggested that a comprehensive analysis of the factor structure was warranted in data from younger patients with spinal deformity, for whom the questionnaire was originally designed, to evaluate whether the same phenomenon was observed,

The aim of this study was to evaluate the factor structure of the English [9, 10], Spanish [7, 8], French [11], Italian [5] and German [12] versions of the SRS-22 outcome instrument in young patients with spinal deformity. We aimed to evaluate whether these versions showed the same four-factor structure for the non-management domains as the original version, and whether this structure was invariant over the different languages.

Materials and Methods

The analyses were carried out on the 20 non-management items of the SRS-22[1] from baseline questionnaire data completed by 3618 young (<20 years) patients with spinal deformity (2713 English-speaking, 270 Spanish, 264 German, 223 Italian, and 148 French). Over 95% had adolescent idiopathic scoliosis, and a small minority had neuromuscular or congenital deformities. Three quarters were female, and their mean (±SD) age was 15.0±2.2 years (Table 1). The English data originated from the British Spine Registry https://www.britishspineregistry.com/. The French and some of the

[1]the official Spanish and Italian versions contained the SRS-22R formulation of item 18, whilst that of SRS-22 was used in all other languages. http://www.srs.org/professionals/online-education-and-resources/patient-outcome-questionnaires

German data came from the idiopathic scoliosis cohorts of the European Spine Study Group's (ESSG) prospective multicentre study https://www.spine-essg.com/. The remaining German data came from the German Spine Society's Spine Registry https://dwg.memdoc.org/. The Italian and Spanish data had been collected in connection with previous observational studies ([7, 8, 13]). IRB approval had been obtained for all the original individual studies.

Statistical analyses

The factor structure of the 20 non-management items of the SRS-22 questionnaire was tested using confirmatory factor analysis (CFA), with structural equation modelling software AMOS 24.0. Confirmatory (rather than exploratory) factor analysis (CFA) is the preferred approach to formally assess a questionnaire's dimensionality when existing theory and empirical evidence support a particular structure [14]. CFA assesses the contribution of each of the questionnaire's questions or "items" ("Item Loading") and measures the adequacy of the measurement model ("Goodness of Fit"). Item loading indicates the strength of the relationship between each item and its underlying factor, with guidelines for minimum values ranging from > 0.4 to >0.7 [15]. Model fit indices reflect the discrepancy between the proposed model structure and the empirical correspondence between model variables and were based on maximum likelihood estimation. The various Goodness of Fit indices included the Root Mean Square Error of Approximation (RMSEA), the ratio of Chi-squared to degrees of freedom ($\chi 2$ /df) and the Comparative Fit Index (CFI). A model is considered to have a good fit if RMSEA is less than 0.05, $\chi 2$ /df is less than 2, and CFI is greater than 0.9 [16].

The analyses comprised systematic comparisons of item-loading and the fit of different models. Equivalence of item-loading was compared across languages for the different

models. The comparability of the language versions was tested by constraining (forcing) item loading to be equal across languages and then testing whether this constrained model was as good a fit to the empirical data as one in which items were allowed to load "freely" (unconstrained) and potentially differ between the languages. The Chi-squared difference test was used to assess the significance of the difference between constrained and unconstrained models; a lack of significant difference confirmed equivalence of item-loading for the different languages.

The following models were evaluated:

Model 1 was a one-factor model that assumed all items loaded on a common factor i.e. “all 20 items measured the same construct”.

Model 2 was the hypothesized four-factor structure with five questions per factor and with item loading estimated for the total sample; in model 2a item-loading was estimated freely for the language subsamples separately.

Model 3 comprised the best fitting item-loading representing all five language versions, i.e. best fit when items were constrained to have the same loading across the different languages.

Model 4 (after confirmation that the 4 worst fitting items were the same as those identified and recommended for removal in Mannion et al 2018 [6]) comprised the four-factor structure for the 16-item version, with the item-loading estimated for the total sample; in model 4a, item-loading was estimated freely for the language subsamples separately.

Model 5 comprised (for the 16-item version) the best fitting item-loading representing all five languages, i.e., best fit when items were constrained to have the same loading across the different languages.

Results

Table 2 shows the fit indicators and results of the 5 CFA models that were tested. The one-factor model (Model 1) did not yield a good fit to the empirical data (RMSEA 0.13, CFI 0.67, Table 2). The originally proposed four-factor structure with dimensions of function, pain, self-image, and mental health, with five items each (Model 2), had a significantly better fit than Model 1 (Chi-squared difference test ($\Delta X^2(7) = 7521.27$, $p < 0.001$, Table 2) and it showed reasonable fit parameters (RMSEA 0.06, CFI 0.92). In model 2a, item-loading was estimated freely and separately for the English, Spanish, German, Italian, and French subsamples, and the fit indices of that model were good (RMSEA = 0.03). However, in Model 3 with the restriction that item loading of each individual question be constrained to be the same across the five languages, there was a significant decrease in the model fit (Model 3 compared with Model 2a: $\Delta X^2(64) = 357.31$, $p < 0.001$). Thus, this model comparison for the 20-item version of the SRS indicated non-equivalence of language versions.

Identical to the findings in patients with adult deformity [6], the items with the lowest loadings on each of the 4 factors in this young population were also Q17 (sick days), Q15 (financial difficulties), Q14 (personal relationships), and Q3 (nervous), with coefficients of 0.49, 0.24, 0.51 and 0.58, respectively (0.41, 0.32, 0.64 and 0.60 in [6]). This justified removal of these same 4 items for examination of the proposed 16-item version of the SRS instrument in Model 4, and the latter achieved a significantly better fit than did the 20-item version (Model 4 compared with Model 2: $\Delta X^2(64) = 1318.12$, $p < 0.001$, Table 2). Some item-loadings in the 16-item solution were relatively low, but still satisfactory by most standards [15] (item 4, 0.55; item 11, 0.55; and item 18, 0.46; Fig 1). In Model 4a, loading was estimated freely and separately for the English, Spanish, German, Italian and French subsamples and the results showed significantly better fit in this trimmed 16-item version compared with the corresponding 20-item instrument (Model 4a compared with Model 2a: $\Delta X^2(320) = 1848.77$, $p < 0.001$, Table

2). However, equivalence across language versions was also not reached in this shorter 16-item SRS: comparison of Model 5 (where the item loading was constrained to be the same across the five language versions) and Model 4a (items loading freely and separately for the different languages) showed a significant decrease of model fit (Model 5 compared with Model 4a: $\Delta X^2(48) = 302.90$, $p < 0.001$, Table 2). Some items showed particularly weak item-loading (≤0.30) for some languages, e.g. item 11 (medication usage) in French (0.25), and item 18 (going out with friends) in German (0.22) (Fig 2).

Discussion

The present study revealed that in young patients with deformity, for whom the SRS instrument was originally devised, the shorter, 16-item version showed a better fit to the instrument's intended 4-factor structure than did the longer 20-item version, just as was found for patients with adult spinal deformity [6]. The study also showed that, although the 4-factor structure was upheld, the fit of the items was not perfect and was not language-invariant. Hence, while the 16-item model was a plausible and acceptable model, and better than the 20-item model, it may not necessarily be the best (conceivable) model, in an absolute sense.

The SRS-22 is the most common outcome instrument used in patients with spinal deformity and has been cross-culturally adapted into many different languages. However, few studies have examined its factor structure or the comparability of the latter across languages [5]. As has been discussed previously [6], the response options for the 20 non-management items of the SRS-22 have many different formats (frequencies, intensities, days, percentages, types of medication, etc), and there is some error specifically associated with this, serving to reduce the comparability between items and the likelihood of (expectedly) similar items loading together on a

common factor. Given this limitation, the four-factor SRS has a surprisingly good fit. When estimates were compared across the five different languages there was no language version that obviously “did not work”, in the sense that its relative item-loading was massively different from the item loadings for the other languages, which was in keeping with the findings in patients with *adult* spinal deformity [6]. Nonetheless, some items showed particularly weak loading in some languages, even in the shortened 16-item instrument. In French, item 11 (medication use) and in German both items 11 and 18 (going out with friends) loaded with coefficients ≤ 0.30. A systematic review of the psychometric properties of all available language versions of the SRS-22 [13] has previously highlighted shortcomings in the French and German translations, reporting negative ratings for most of the investigated psychometric properties and even going so far as to suggest that different questionnaires may be required to assess health-related quality of life in patients with scoliosis in these countries. We are not convinced that these language versions are entirely without worth, and suggest that the problem may instead relate to the wording of specific questions. Items 11 and 18 have been highlighted as problematic in other cross-cultural adaptation studies [7, 8, 17] where they failed to load on the expected domains or had poor internal consistency. The wording of item 18 was subsequently improved in what is now known as the SRS-22R version [10]. In the present study, the improved wording of item 18 had been used in the Spanish and Italian versions of the instrument (the languages for which this item also had the highest factor loadings), but not in the other languages. Unfortunately, revision of the English version of the SRS-22 to produce the SRS-22R was not accompanied by a simultaneous revision of the problematic items in all other languages. In the original studies that provided the data for the present study, if no SRS-22R (revised) version for the given language was available in the literature at the start of data collection (as

was the case for French, German and English), then the wording of the items from the original SRS-22 (or its predecessor, the SRS-30) was used. We recommend that the wording of item 18 be consistently changed to that of the SRS-22R, for all languages, and that this version be used exclusively in future.

The question on pain medication (item 11) also presents numerous issues. Reference is not made specifically to "pain" medication (just "medication usage for your back") in most languages other than English, perhaps explaining its generally low loading on the pain domain in the other language versions. Other issues concern the complexity of the language used in the response options ("strong/weak painkillers" versus "narcotics/non-narcotics") and the lack of specific examples in terms of common trade-names (e.g. none given in the Spanish version) that might otherwise allow the patient to readily identify the category to which their own medication belongs. Other language-specific issues concern item 4, which does not include the word "shape" (in enquiring how one would feel about spending the rest of one's life with the back (shape) one has now) in the official Spanish version of the SRS-22R (http://www.srs.org/UserFiles/file/outcomes/srs-22_spanish.pdf). This item showed a somewhat lesser fit in the Spanish data, compared with the other languages, and the discrepancy should be amended in the official Spanish version of the instrument to improve comparability across languages.

In their systematic review, Monticone et al [13] advised that, following confirmatory factor analyses to verify the structure of the cross-cultural adaptations of the SRS-22, more extensive studies of their psychometric properties (e.g., responsiveness) should be carried out. The present study suggests that this should perhaps be done using the shorter, more structurally valid version of the instrument shown in Fig 1, in both adult and adolescent patient groups. In the present study, the data were extracted from existing databases rather than collected prospectively for the given

research question. Moreover, approximately 5% patients (unable to be specifically identified within the registry data, but estimated from the surgeons' knowledge of the case mix) did not have adolescent idiopathic scoliosis, but instead some type of neuromuscular or congenital deformity. We recommend that further studies be carried out of the test-retest reliability and responsiveness of this shorter version as a stand-alone instrument in each language, in young patients with spinal deformity — and specifically in a homogeneous group of patients with AIS — to ensure that the reduction in the number of items does not threaten other psychometric properties of the instrument.

Questionnaire completion can be a burden for patients, especially those who are involved in long-term, prospective studies with repeated requests for questionnaire completion over time. If a shorter, more structurally valid version of an instrument is available, it should be implemented. This would also have the benefit of reducing the burden on, and costs to, the research team that has to administer the questionnaire, check its completeness, and input data. The use of one and the same, improved, 16-item version of the SRS instrument, in both adult and adolescent deformity patients alike, offers the further advantage of easing the administrative load when considering the routine outcome assessment of all deformity patients within a given spine center.

In summary, use of the shorter, more structurally valid version of the SRS-instrument, with removal of ill-fitting items and adaptation and standardization of the other items across language versions, is expected to deliver more meaningful information on patient-reported outcomes whilst reducing the burden on patients and researchers.

References

1. Jain A, Sponseller PD, Negrini S, Newton PO, Cahill PJ, Bastrom TP, Marks MC, Harms Study G (2015) SRS-7: A Valid, Responsive, Linear, and Unidimensional Functional Outcome Measure for Operatively Treated Patients With AIS. Spine (Phila Pa 1976) 40:650-655. doi: 10.1097/BRS.0000000000000836
2. Caronni A, Zaina F, Negrini S (2014) 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. Research in Developmental Disabilities 35:784-799
3. Rothenfluh DA, Neubauer G, Klasen J, Min K (2012) Analysis of internal construct validity of the SRS-24 questionnaire. Eur Spine J 21:1590-1595. doi: 10.1007/s00586-012-2169-3
4. Jain A, Lafage V, Kelly MP, Hassanzadeh H, Neuman BJ, Sciubba DM, Bess S, Shaffrey CI, Ames CP, Scheer JK, Burton D, Gupta MC, Hart R, Hostin RA, Kebaish KM, International Spine Study G (2016) Validity, Reliability, and Responsiveness of SRS-7 as an Outcomes Assessment Instrument for Operatively Treated Patients With Adult Spinal Deformity. Spine (Phila Pa 1976) 41:1463-1468. doi: 10.1097/BRS.0000000000001540
5. Monticone M, Baiardi P, Calabro D, Calabro F, Foti C (2010) Development of the Italian Version of the Revised Scoliosis Research Society-22 Patient Questionnaire, SRS-22r-I. SPINE 35:1412-1417
6. Mannion AF, Elfering A, Bago J, Pellise F, Vila-Casademunt A, Richner-Wunderlin S, Domingo-Sabat M, Obeid I, Acaroglu E, Alanay A, Perez-Grueso FS, Baldus CR, Carreon LY, Bridwell KH, Glassman SD, Kleinstuck F, European Spine Study G (2018) Factor analysis of the SRS-22 outcome assessment instrument in patients with adult spinal deformity. Eur Spine J 27:685-699. doi: 10.1007/s00586-017-5279-0
7. Bago J, Climent JM, Ey A, Perez-Grueso FJ, Izquierdo E (2004) The Spanish version of the SRS-22 patient questionnaire for idiopathic scoliosis: transcultural adaptation and reliability analysis. Spine (Phila Pa 1976) 29:1676-1680
8. Climent JM, Bago J, Ey A, Perez-Grueso FJ, Izquierdo E (2005) Validity of the Spanish version of the Scoliosis Research Society-22 (SRS-22) Patient Questionnaire. Spine (Phila Pa 1976) 30:705-709
9. Asher MA, Min Lai S, Burton DC (2000) Further development and validation of the Scoliosis Research Society (SRS) outcomes instrument. Spine (Phila Pa 1976) 25:2381-2386
10. Asher MA, Lai SM, Glattes RC, Burton DC, Alanay A, Bago J (2006) Refinement of the SRS-22 Health-Related Quality of Life questionnaire Function domain. Spine (Phila Pa 1976) 31:593-597. doi: 10.1097/01.brs.0000201331.50597.ea 00007632-200603010-00018 [pii]
11. Beauséjour M, Joncas J, Goulet L, Roy-Beaudry M, Parent S, Grimard G, Forcier M, Lauriault S, Labelle H (2009) Reliability and Validity of Adapted French Canadian Version of Scoliosis Research Society Outcomes Questionnaire (SRS-22) in Quebec. SPINE 34:623-628
12. Niemeyer T, Schubert C, Halm HF, Herberts T, Leichtle C, Gesicki M (2009) Validity and reliability of an adapted german version of scoliosis research society-22 questionnaire. Spine (Phila Pa 1976) 34:818-821. doi: 10.1097/BRS.0b013e31819b33be
13. Monticone M, Nava C, Leggero V, Rocca B, Salvaderi S, Ferrante S, Ambrosini E (2015) Measurement properties of translated versions of the Scoliosis Research

Society-22 Patient Questionnaire, SRS-22: a systematic review. Qual Life Res 24:1981-1998. doi: 10.1007/s11136-015-0935-5
14. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC (2010) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 10:22. doi: 1471-2288-10-22 [pii]
10.1186/1471-2288-10-22
15. Knekta E, Runyon C, Eddy S (2019) One Size Doesn't Fit All: Using Factor Analysis to Gather Validity Evidence When Using Surveys in Your Research. CBE Life Sci Educ 18:rm1. doi: 10.1187/cbe.18-04-0064
16. Schermelleh-Engel K, Moosbrugger H, Müller H (2003) Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. . Methods of Psychological Research Online 8:23–74
17. Alanay A, Cil A, Berk H, Acaroglu E, Yazici M, Akcali O, Kosay C, Genc Y, Surat A (2005) Reliability and Validity of Adapted Turkish Version of Scoliosis Research Society-22 (SRS-22) Questionnaire. SPINE 30:2464-2468

Table 1. Demographic and baseline symptom severity data (SRS subtotal score) of the patients

	N	Age (years)	Gender (F:M: %F)	SRS subtotal score (20-item)
English	2713	15.0 ± 2.0	2044:668 (75.4)[a]	3.6 ± 0.6
Spanish	270	14.4 ± 1.9	236:34 (87.4)	3.9 ± 0.5
German	264	15.2 ± 1.8	212:52 (80.3)	3.7 ± 0.5
Italian	223	14.1 ± 2.0	120:103 (53.8)	4.1 ± 0.5
French	148	15.0 ± 2.0	118:27 (81.4)*	3.9 ± 0.6
ALL PATIENTS	3618	15.0 ± 2.0	2730:884 (75.5)	3.6 ± 0.6

[a] some missing data for gender in 1 of the English and 3 of the French datasets

Table 2. Fit indicators and results of the 5 confirmatory factor analysis (CFA) models that were tested

	N	X^2	df	X^2 / df	RMSEA	CFI	Δ X^2 (df)	p
Model 1 “All the same” –one-factorial model	3618	10074.57	169	59.61	.13	.67		
Model 2 “4 Factors” (20 items with 5 items each on Function, Pain, Self-Image, and Mental Health)	3618	2553.30	162	15.76	.06	.92	Model 2 better than Model 1? Yes: 7521.27 (7)	<.001
Model 2a “as model 2 with items estimated freely for English (2713), Spanish (270), German (264), Italian (223), and French (148) samples"	3618	3703.36	810	4.57	.03	.90		
Model 3 “as model 2 with items constrained to have the same loading in English (2713), Spanish (270), German (264), Italian (223), and French (148) samples"	3618	4070.67	874	4.66	.03	.89	Model 3 as good as Model 2a? No, it is worse: 357.31 (64)	<.001
Model 4 "Trimmed model (16 item model with just 4 items each for Function, Pain, Self-Image, Mental health with exclusion of items 3, 14, 15, 17)"	3618	1235.18	98	12.60	.06	.96	Model 4 better than Model 2? Yes: 1318.12 (64)	<.001
Model 4a “as model 4 with items estimated freely for English (2713), Spanish (270), German (264), Italian (223), and French (148) samples”	3618	1854.59	490	3.78	.03	.95	Model 4a better than Model 2a? Yes: 1848.77 (320)	<.001
Model 5 "as model 4 with items constrained to have the same loading in English (2713), Spanish (270), German (264), Italian (223), and French (148) samples"	3618	2157.49	538	4.01	.03	.94	Model 5 as good as Model 4a? No: 302.90 (48)	<.001

Note. $\chi 2$ = Chi-square value indicates the minimum discrepancy between empirical covariance structures and those implied by the model; df = Degrees of freedom; $\chi 2$/df = Minimum discrepancy divided by its degrees of freedom, as an indicator of fit; *p* = p-value of Minimum discrepancy divided by its degrees of freedom, which should be smaller than 2 ([16]); CFI = Comparative fit index; CFI higher than .90 in the mediation model reflect acceptable fit between the model and the data ([16]). RMSEA value below .05 reflects a good fit of the model ([16]). To test the fit between two nested models the difference in $\chi 2$ and df Δ $X^{2(}$df) was calculated ($\chi 2$ difference test). P indicates a significantly better fit of the model with lower $\chi 2$ value.