

Dimensionality, Reliability, and Validity of the Revised Fibromyalgia Impact Questionnaire in Two Spanish Samples

JUAN V. LUCIANO,¹ JAUME AGUADO,² ANTONI SERRANO-BLANCO,³ ELENA P. CALANDRE,⁴ AND CARMEN M. RODRIGUEZ-LOPEZ⁴

Objective. The present study attempted to fill a research gap by performing the first dimensionality analysis of the Revised Fibromyalgia Impact Questionnaire (FIQR) using exploratory and confirmatory techniques. A second objective was to report on the reliability and construct validity of the FIQR in Spanish patients.

Methods. FIQR data from a sample of adult fibromyalgia patients (n = 113) were analyzed using principal components analysis (PCA). Subsequently, a set of confirmatory factor analyses (CFAs) was conducted in another sample (n = 179) to analyze the goodness of fit of various factor models. FIQR reliability was assessed by computing Cronbach's alpha and coefficient H. Construct validity was evaluated by comparing the FIQR scores of participants categorized by employment status.

Results. According to the PCA, the FIQR structure might be described as having 1 global factor of functional impairment. Although subsequent CFAs confirmed that 1 factor accounted for the greatest proportion of common variance in the FIQR items, a confirmatory bifactor analysis indicated that the items were multidimensional because of their simultaneous significant loading on specific factors. The Cronbach's alpha values of the FIQR domains were very good (>0.80) and the H estimate for the FIQR total score was excellent (0.93). Overall, the FIQR domains were able to distinguish between patients differing in employment status (working outside the home versus on sick leave).

Conclusion. Our results indicate that the Spanish version of the FIQR has a complex factor structure, has excellent reliability, and shows good construct validity.

INTRODUCTION

Fibromyalgia (FM) is a prevalent, debilitating syndrome of an unknown etiology that is mainly characterized by chronic widespread pain, fatigue, disturbed sleep, and psychological distress (1,2). The Outcome Measures in

Rheumatology 9 initiative (3) identified a set of 9 symptom domains to be assessed in FM treatment trials (physical function, patient global impression of change, pain intensity, fatigue, cognitive dysfunction, multidimensional function/health-related quality of life [HRQOL], sleep disturbance, tenderness, and depression). According to a recent review (4), the Revised Fibromyalgia Impact Questionnaire (FIQR) (5) might be considered the gold standard instrument for assessing physical function (FIQR physical function domain), patient global improvement, and multidimensional function/HRQOL (FIQR total score).

The FIQ (6), a previous version of the FIQR, is the most extensively used assessment tool in FM (7); it has been translated into 14 languages and cited in over 300 articles. However, the FIQ has some issues related to wording, omissions, concepts, and scoring that need to be resolved to improve instrument efficiency (5). For example, the FIQ uses a visual analog scale that requires patients to slash a 100-mm line and is scored with a ruler, and scoring is complex because scores on the 11 physical function items are added together, then divided by the number of items answered, and finally multiplied by 3.33 to yield a 0–10 composite physical function score. The “felt good” “day-

Dr. Luciano's work was supported by the Institute of Health Carlos III (grant Red RD06/0018/0017).

¹Juan V. Luciano, PhD: Parc Sanitari Sant Joan de Déu, Sant Boi del Llobregat, Primary Care Prevention and Health Promotion Research Network, Madrid, and Open University of Catalonia, Barcelona, Spain; ²Jaume Aguado, MSc: Parc Sanitari Sant Joan de Déu, Sant Boi del Llobregat, Spain; ³Antoni Serrano-Blanco, MD, PhD: Parc Sanitari Sant Joan de Déu, Sant Boi del Llobregat, and Primary Care Prevention and Health Promotion Research Network, Madrid, Spain; ⁴Elena P. Calandre, MD, Carmen M. Rodriguez-Lopez, MD: Instituto de Neurociencias Federico Oloritz, University of Granada, Granada, Spain.

Address correspondence to Juan V. Luciano, PhD, Research and Development Unit, Parc Sanitari Sant Joan de Déu, Calle Doctor Antoni Pujadas 42, 08830, Sant Boi de Llobregat, Barcelona, Spain. E-mail: jvluciano@pssjd.org.

Submitted for publication January 1, 2013; accepted in revised form April 11, 2013.

Significance & Innovations

- The present study has furthered our understanding of the latent structure of the Revised Fibromyalgia Impact Questionnaire (FIQR) through the comparison of viable factor models for the instrument using confirmatory factor analysis.
- Besides reporting the total score when using the FIQR, a confirmatory bifactor analysis satisfied the need to compute and report the scores on the 3 FIQR domains.
- The FIQR domains showed excellent reliability and construct validity in Spanish fibromyalgia patients.

of-the-week” item is reverse scored and the result is multiplied by 1.43 to yield a 0–10 score, whereas the “missed work” “day-of-the-week” score is generated by multiplying the number of days by 1.43 to yield a 0–10 score. The cross-cultural validity of the physical impairment items is questionable because they collect data on activities or actions typical of women living in developed countries. Furthermore, relevant symptoms in FM, such as cognitive dysfunction, tenderness, balance, and environmental sensitivity, are not assessed in the FIQ.

The FIQR was developed to address the problems listed above (5). The FIQR includes 21 individual items that are all answered on an 11-point numerical rating scale (range 0–10), where 10 reflects greater impairment. The timeframe is the previous 7 days, and the items are distributed into 3 associated domains: “function” (9 items), “overall impact” (2 items that now address the overall impact of FM on functioning and the overall impact on symptom severity), and “severity of symptoms” (10 items). In the FIQR, the third domain includes 4 new items related to memory, tenderness, balance, and sensitivity to loud noises, bright lights, odors, and cold temperatures. One practical modification is that the scoring system is much easier; the physical function domain (range 0–90) is divided by 3, the overall impact domain (range 0–20) is not transformed, and the severity of symptoms domain (range 0–100) is divided by 2. The FIQR total score (range 0–100) is obtained by adding the 3 domain scores. The FIQR has been shown to be a psychometrically sound instrument, is equivalent to the original version (the FIQ and FIQR are strongly correlated, which makes it possible to compare studies), and is clinically useful because it can be completed by patients in <2 minutes and scored in ~1 minute (5).

Ediz and colleagues (8) assessed the reliability and validity of the FIQR in a Turkish FM sample. The authors not only confirmed the excellent internal consistency of the FIQR, with Cronbach’s alpha values of ~0.90 in both assessment periods, but also demonstrated good stability of the FIQR total score over time ($r = 0.83$). More recently, Srifi et al (9) adapted the FIQR to the Moroccan linguistic and cultural context to assess its psychometric properties in a sample of 80 patients with FM recruited in

a hospital. Once again, the internal consistency was very good ($\alpha \geq 0.90$ in both study periods), as was test–retest reliability ($r = 0.84$).

Although the psychometric properties of the FIQR have been extensively tested, the current lack of factor analytical studies of the instrument is surprising. In our opinion, it is crucial to discover whether there is a statistical basis for combining the 21 FIQR items into 3 different domains as well as into a single total score (5). Consequently, the present study had the 2 following interrelated objectives: 1) we addressed the abovementioned research gap by examining the dimensionality of the Spanish version of the FIQR in 2 independent samples of FM patients using both exploratory and confirmatory factor analytical procedures, and 2) the reliability and construct validity of the FIQR dimensions were also examined.

MATERIALS AND METHODS

Participants. A detailed description of the participants was previously provided (ref. 10 and Salgueiro M, et al: unpublished observations). Sample 1 included 113 patients with FM who participated in a study aiming to adapt the FIQR to Spanish-speaking patients (Salgueiro M, et al: unpublished observations). Sample 2 included 179 patients who anonymously took part in a survey designed to investigate the prevalence of previous suicide attempts among people with FM (10). The participants were recruited from various Spanish FM patient associations. Both studies were approved by the University of Granada (Spain) Ethics Committee and performed in compliance with the Declaration of Helsinki and its subsequent updates. The sociodemographic data and FIQR total score of each sample are shown in Table 1.

Measures. Patients from both study samples completed a sociodemographic questionnaire and the FIQR (5) as part of a paper-and-pencil battery of instruments. The process of adapting the FIQR to Spanish and preliminary evidence of the psychometric properties of the FIQR in Spanish-speaking FM patients were recently provided by Salgueiro et al (unpublished observations), who found adequate test–retest reliability (1 week) using Spearman’s rank correlations (FIQR total $r = 0.81$, FIQR function $r = 0.77$, FIQR overall impact $r = 0.51$, and FIQR severity of symptoms $r = 0.83$) and good convergent validity with the Hospital Anxiety and Depression Scale (HADS) anxiety subscale ($r = 0.67$), the HADS depression subscale ($r = 0.68$), the Brief Pain Inventory severity index ($r = 0.68$), and the Brief Pain Inventory interference index ($r = 0.85$). Moreover, the correlations between the FIQR total score and the Short Form 36 subscales were all statistically significant and large ($r \geq 0.50$) in most cases. The pattern of relationships between the FIQR domains and the aforementioned instruments was also reported by Salgueiro et al (unpublished observations).

Statistical analyses. SPSS, version 19.0 and Mplus, version 7.0 were used to conduct the statistical analyses.

Table 1. Sociodemographic and FIQR data in samples 1 and 2*

	Sample 1 (n = 113)†	Sample 2 (n = 179)†	P
Women, no. (%)	109 (88.5)	175 (97.8)	0.75
Age, mean \pm SD years	51.5 \pm 9.6	51 \pm 8.5	0.64
Education level, no. (%)			0.24
No school	9 (8.2)	14 (8.0)	
Primary school	50 (45.5)	99 (56.6)	
Secondary school	37 (33.6)	41 (23.4)	
University	14 (12.7)	21 (12.0)	
Marital status, no. (%)			0.01
Single	23 (21.9)	11 (8.6)	
Married/de facto	66 (62.9)	100 (78.1)	
Divorced	13 (12.4)	10 (7.8)	
Widowed	3 (2.9)	7 (5.5)	
Employment status, no. (%)			0.05
Work only at home	20 (22.7)	47 (26.7)	
Work outside the home	20 (22.7)	34 (19.3)	
Unemployed	22 (25.0)	24 (13.6)	
Sick leave	13 (14.8)	49 (27.8)	
Retired	13 (14.8)	22 (12.5)	
Years since FM diagnosis, mean \pm SD	8.5 \pm 7.7	7.1 \pm 4.8	0.07
FIQR total score, mean \pm SD (range 0–100)	68.2 \pm 17.5	66.2 \pm 20.7	0.41
FIQR function, mean \pm SD (range 0–30)	18.9 \pm 6.7	18.7 \pm 7.2	0.97
FIQR overall impact, mean \pm SD (range 0–20)	11.8 \pm 5.6	11.7 \pm 6.2	0.91
FIQR severity of symptoms, mean \pm SD (range 0–50)	37.5 \pm 8.7	35.8 \pm 10.5	0.09
* FIQR = Revised Fibromyalgia Impact Questionnaire; FM = fibromyalgia.			
† Not all patients answered every section of the sociodemographic questionnaire; therefore, the total number of responders is not the same in all variables.			

Dimensionality. Following a cross-validation approach, the 2 FM samples were used to examine the FIQR factor structure. We performed a principal components analysis (PCA) with the first sample and confirmatory factor analyses (CFAs) with the second sample. The sample size for both the PCA and CFAs was adequate because we were able to satisfy the recommendation of a minimum of 5 participants per item (11).

First, using sample 1, responses to the 21 FIQR items were subjected to PCA with oblique (Oblimin) rotation. The suitability of data for factor analysis was examined by means of Kaiser-Mayer-Olkin's measure of sampling adequacy (KMO) (12). Bartlett's test of sphericity (13) was also applied to examine the extent to which the correlation matrices departed from orthogonality. We used the following combination of rules to determine the optimal number of components to retain (14): Kaiser's criterion (components with eigenvalues >1.0), the ratio of the eigenvalue of the first and second unrotated component (a ratio >3.0 suggests unidimensionality), Cattell's scree test, and item loadings (an item forms part of a factor if its factor loading on that factor is ≥ 0.40).

Second, using sample 2, a set of CFAs was performed. The maximum likelihood estimation with robust SEs was applied to test the fit of the different factor models. Although a model with a nonsignificant chi-square estimate is generally considered a good-fitting model, Hu and Bentler (15) recommended combination rules to evaluate model fit. The following indices were analyzed (values in parentheses denote goodness-of-fit standards): the Tucker-Lewis index (TLI; ≥ 0.95), comparative fit index (CFI;

≥ 0.95), and root mean square error of approximation (RMSEA) with its 90% confidence interval (≤ 0.06). Used together, these indices provide a more conservative and reliable evaluation of the factorial solution as well as different information about model fit.

Reliability. Internal consistency was assessed by means of Cronbach's alpha, which reflects the average intercorrelation among all items. The rule of thumb for describing internal consistency is as follows: $\alpha \geq 0.70$ is defined as acceptable, $\alpha \geq 0.80$ is defined as good, and $\alpha \geq 0.90$ is defined as excellent (11). The homogeneity of the FIQR was also evaluated on the basis of the corrected item-total correlations.

Construct validity. We used a known-groups validity approach, which is founded on the basis that certain specified groups of patients are expected to score differently from others. Student's *t*-test for independent samples (with unequal variances) was performed to assess the construct validity of the FIQR and its 3 domains for discriminating between the patients that were on sick leave (objective indicator of disability) and those that were working outside the home. It was hypothesized that the former would have worse (or higher) functional impairment than the latter. The alpha level was set at *P* less than 0.05.

RESULTS

We examined differences in sociodemographic characteristics and FIQR scores between samples 1 and 2, applying

Table 2. Means and SDs, corrected item-total correlations (r_{tot}), and factor loadings (λ) for all FIQR items in samples 1 and 2*

FIQR items, Spanish version†	Sample 1 (n = 113)				Sample 2 (n = 179)					
	Mean \pm SD	r_{tot}	PCA λ (1F)		Mean \pm SD	r_{tot}	CFAs, λ (1F)	CFAs, λ (3F)‡	CBFA, global	CBFA, specific
Function										
1. Peinarse (Brush or comb your hair)	4.39 \pm 2.89	0.67	0.71		4.17 \pm 3.09	0.65	0.72	0.65	0.73	−0.12§
2. Caminar 20 minutos sin necesidad de pararse (Walk continuously for 20 minutes)	5.61 \pm 3.32	0.74	0.78		5.80 \pm 3.39	0.70	0.77	0.73	0.77	−0.03§
3. Preparar la comida (Prepare a homemade meal)	4.54 \pm 3.03	0.63	0.67		5.01 \pm 3.10	0.75	0.79	0.81	0.81	0.11§
4. Barrer, fregar, o pasar la aspiradora (Vacuum, scrub, or sweep floors)	6.94 \pm 2.72	0.69	0.73		7.09 \pm 2.87	0.70	0.71	0.80	0.71	0.53
5. Levantar y transportar una bolsa de la compra llena (Lift and carry a bag full of groceries)	7.72 \pm 2.49	0.57	0.63		7.83 \pm 2.64	0.72	0.73	0.80	0.70	0.55
6. Subir escaleras (Climb 1 flight of stairs)	7.13 \pm 2.61	0.65	0.70		7.30 \pm 2.86	0.77	0.81	0.80	0.79	0.16
7. Cambiar la ropa de la cama (Change bed sheets)	6.73 \pm 2.47	0.58	0.63		6.44 \pm 3.18	0.72	0.74	0.81	0.73	0.34
8. Estar sentada en una silla durante 45 minutos (Sit in a chair for 45 minutes)	6.80 \pm 2.75	0.52	0.58		6.98 \pm 3.07	0.65	0.68	0.66	0.67	0.04§
9. Hacer la compra (Go shopping for groceries)	6.37 \pm 2.73	0.66	0.71		6.58 \pm 3.08	0.70	0.73	0.77	0.74	0.15§
Overall impact										0.74¶
10. La fibromialgia me impidió hacer lo que tenía proyectado esta semana (Fibromyalgia prevented me from accomplishing goals for the week)	5.89 \pm 2.85	0.51	0.54		5.72 \pm 3.25	0.53	0.52	0.91	0.53	
11. Los síntomas de mi fibromialgia me tuvieron totalmente abrumada (I was completely overwhelmed by my fibromyalgia symptoms)	5.94 \pm 3.24	0.45	0.49		6.01 \pm 3.38	0.52	0.51	0.89	0.51	
Severity of symptoms										
12. Dolor (Pain)	7.67 \pm 2.18	0.69	0.73		7.68 \pm 2.08	0.69	0.67	0.72	0.65	0.31
13. Energía (Energy)	7.76 \pm 2.13	0.65	0.70		7.20 \pm 2.72	0.53	0.52	0.58	0.49	0.31
14. Rigidez (Stiffness)	7.37 \pm 2.22	0.60	0.65		6.90 \pm 2.77	0.71	0.69	0.74	0.66	0.34
15. Calidad del sueño (Quality of your sleep)	8.57 \pm 2.13	0.48	0.53		8.14 \pm 2.73	0.66	0.61	0.74	0.56	0.51
16. Depresión (Depression)	6.45 \pm 3.21	0.63	0.68		6.19 \pm 3.15	0.72	0.70	0.76	0.67	0.37
17. Problemas de memoria (Memory problems)	7.46 \pm 2.47	0.55	0.61		7.20 \pm 2.77	0.59	0.53	0.65	0.49	0.46
18. Ansiedad (Anxiety)	7.07 \pm 2.78	0.68	0.72		7.06 \pm 3.07	0.70	0.65	0.77	0.60	0.52
19. Dolorimiento al tacto (Tenderness to touch)	7.39 \pm 2.75	0.70	0.74		7.40 \pm 2.76	0.76	0.76	0.82	0.72	0.36
20. Problemas de equilibrio (Balance problems)	6.90 \pm 2.47	0.60	0.64		6.31 \pm 3.06	0.68	0.67	0.76	0.63	0.44
21. Grado de sensibilidad al ruido intenso, la luz brillante, los olores, el frío (Sensitivity to loud noises, bright lights, odors, and cold)	8.04 \pm 2.22	0.50	0.54		7.47 \pm 2.77	0.70	0.70	0.72	0.68	0.26

* FIQR = Revised Fibromyalgia Impact Questionnaire; PCA = principal components analysis; 1F = 1-factor model; CFAs = confirmatory factor analyses; 3F = 3-factor model; CBFA = confirmatory bifactor analysis.
† The original wording of items in English is shown in parentheses.
‡ Correlation among FIQR factors: F1 and F2 = 0.53, F1 and F3 = 0.83, and F2 and F3 = 0.50.
§ Nonsignificant loadings in the specific factor.
¶ A correlated error (not a specific factor) was modeled for items 10 and 11 (domain 2) because a latent factor with only 2 indicators would not be identified in a bifactor model. The value shown in the table represents the common residual variance shared between the items.

Student's *t*-test for continuous variables and the chi-square test for categorical variables (Table 1). There were statistically significant differences ($P < 0.05$) in marital status and employment status between the samples. The FIQR total scores were not significantly different, varying from 15.5–100 (mean \pm SD 68.2 \pm 17.5) in sample 1 and from 10.7–98.2 (mean \pm SD 66.2 \pm 20.7) in sample 2.

Descriptive statistics. Descriptive statistics were computed for all FIQR items as shown in Table 2. Each item was examined in terms of the mean, SD, and corrected item-total correlations. All items obtained a corrected item-total correlation that was higher than the rule of thumb minimum value of 0.20 (16). The corrected item-

total correlation coefficients ranged from 0.45 (item 11) to 0.74 (item 2) in sample 1 and from 0.52 (item 11) to 0.77 (item 6) in sample 2. Given the absence of univariate and multivariate outliers, all cases were retained for the following statistical analyses.

PCA. The KMO measure produced a coefficient of 0.88, which is indicative of satisfactory sampling adequacy. Bartlett's test of sphericity produced a figure of 1,330.47 ($P < 0.0001$), indicating that the correlation matrix was unlikely to be an identity matrix and was therefore suitable for factor analysis. The PCA ($n = 109$ after listwise deletion) revealed 5 factors with eigenvalues > 1.0 . The principal factor accounted for 43.2% of total variance,

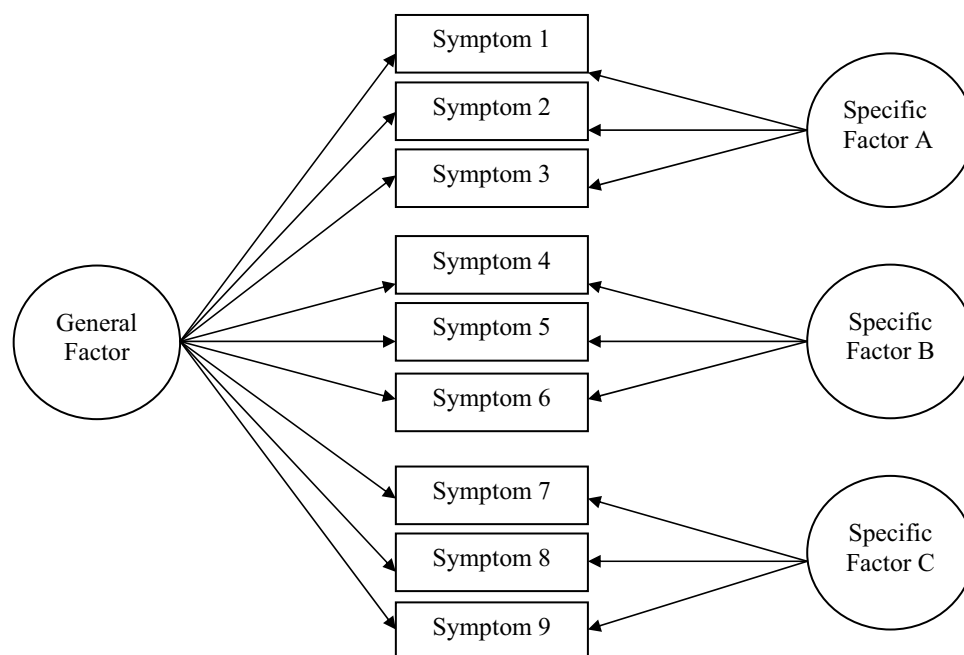


Figure 1. Example of a bifactor structural model.

whereas the other 4 factors accounted for 8.6%, 7%, 5.6%, and 4.8% of variance (the eigenvalues of the 5 factors were 9.06, 1.80, 1.46, 1.19, and 1.01, respectively). Given that the criterion of eigenvalues >1.0 can lead to overestimating the number of meaningful factors, the ratio of the first to the second eigenvalue was >3.0 , and the inspection of the scree plot of the eigenvalues suggested that 1 factor may be sufficient, a second PCA was carried out, specifying that only 1 factor should be identified. The second PCA yielded a 1-factor solution, with the 21 items loading strongly on the factor (Table 2).

CFAs. A sequence of the following 5 models was tested in sample 2 ($n = 156$ after listwise deletion): model 1 = a single factor as found with the PCA on which all the FIQR items load; model 2 (respecification of model 1) = a single factor incorporating correlated residuals (items situated in the same instrument domain are likely to covary and, as a consequence, 82 correlated residuals were specified); model 3 (respecification of model 2) = corresponds to model 2 but dropped the statistically nonsignificant correlated residuals; and model 4 = a 3-factor model reflecting the presence of 3 interrelated domains. Models 1–4 were lower-order models that did not provide an opportunity to study the hierarchical structure of the FIQR; therefore, we also computed model 5 (confirmatory bifactor analysis [CBFA]). From this approach (17), it is posited that all items are saturated with a general latent factor and 3 specific factors that are mutually uncorrelated, vary independently of the general factor, and account for unique variance among the items beyond the general factor. We decided to estimate a bifactor model because it could help to confirm whether the FIQR items are multidimensional, which would justify the computation of subscale scores, or whether the items are mainly unidimensional and only 1 total FIQR score should be reported. For illustrative

purposes, a generic example of a bifactor structural model is shown in Figure 1.

Fit statistics for the 5 factor models are shown in Table 3. According to fit criteria, model 1 did not represent the observed data well. As expected, an inspection of localized areas of strain in model 1 indicated that there was evidence of correlated residuals among items situated in the same instrument domain (e.g., items 10 and 11; modification index = 68.1 and standardized expected parameter change = 0.77). The inclusion of the correlated residuals in model 2 yielded an excellent-fitting solution, superior to the 1-factor model without residual covariations. However, 50 of the correlated residuals were not statistically significant and were removed in model 3. This model fitted the data almost identically to model 2 but was chosen as the best-fitting 1-factor model because of parsimony considerations. The 32 correlated residuals that were statistically significant in the final 1-factor model ranged from a minimum of 0.12 ($\theta_{18,21}$) to a maximum of 0.74 ($\theta_{10,11}$). Concerning the 3-factor model (model 4) and the CBFA (model 5), although these models did not reach the conservative rules of thumb applied to the CFI/TLI and RMSEA indices, they were considered good approximations to the data. In the 3-factor model, the correlations among the FIQR factors were all significant ($P < 0.01$) and large ($r \geq 0.50$). In the CBFA, the results suggested that the 3 specific factors reflected meaningful residual variance not accounted for by the general factor.

The standardized factor loadings were all statistically significant ($P < 0.01$) and large ($\lambda > 0.50$) in the final 1-factor model (ranging from a minimum of 0.51 [item 11] to a maximum of 0.81 [item 6]) as well as in the 3-factor model (ranging from a minimum of 0.58 [item 13] to a maximum of 0.91 [item 10]). The results for the specific factors in the CBFA were variable. Five items (items 1, 2, 3, 8, and 9) did not have statistically significant factor load-

Table 3. Fit statistics for the factor models of the FIQR in sample 2 (n = 179)*

Model	Source	χ^2	df	TLI	CFI	RMSEA (90% CI)
1	1-factor	515.9	189	0.78	0.80	0.11 (0.09–0.12)
2	1-factor (+ θ_{1-9} , θ_{10-11} , θ_{12-21} free)†	138.2‡	107	0.96	0.98‡	0.04 (0.02–0.06)‡
3	1-factor (+ statistically significant θ free)§	208.1	157	0.97‡	0.96	0.05 (0.03–0.06)
4	3-factor (3 intercorrelated domains)	323.1	186	0.92	0.91	0.07 (0.06–0.08)
5	Bifactor (1 general factor + 3 uncorrelated domains)	265.7	169	0.94	0.93	0.06 (0.05–0.07)

* FIQR = Revised Fibromyalgia Impact Questionnaire; TLI = Tucker-Lewis index; CFI = comparative fit index; RMSEA = root mean square error of approximation; 90% CI = 90% confidence interval of the RMSEA.
† Correlated residuals among items included in FIQR section 1 (items 1–9), section 2 (items 10–11), and section 3 (items 12–21).
‡ Indices that indicate the best model fit.
§ The 32 statistically significant correlated residuals are $\theta_{1,5-7}$, $\theta_{2,8}$, $\theta_{3,4}$, $\theta_{3,7}$, $\theta_{4,5}$, $\theta_{4,7}$, $\theta_{5,6-7}$, $\theta_{7,9}$, $\theta_{8,9}$, $\theta_{10,11}$, $\theta_{12,14}$, $\theta_{12,19}$, $\theta_{13,14-15}$, $\theta_{14,15}$, $\theta_{15,16-18}$, $\theta_{15,21}$, $\theta_{16,17-18}$, $\theta_{16,20}$, $\theta_{17,18}$, $\theta_{18,20-21}$, $\theta_{18,19-21}$, $\theta_{19,20}$.

ings on the physical specific factor; that is, this specific factor was reduced to 4 significant items (median loading 0.40) after controlling for variance due to the general factor. In contrast, the symptoms specific factor included moderate and consistent loadings of its 10 items (median loading 0.39). In the general latent factor, all factor loadings were statistically significant and ranged from 0.49 (items 13 and 17) to 0.81 (item 3), with an average of 0.66. We decided to retain the 3 FIQR domains for further analyses (reliability and validity), given the bifactor model showed a reasonable fit to the data.

Reliability. The overall Cronbach's alpha coefficient was excellent in sample 1 (0.93) as well as in sample 2 (0.95). Concerning the 3 FIQR domains, the Cronbach's alpha coefficients ranged from good to excellent in sample 1 (function = 0.90, overall impact = 0.81, and severity of symptoms = 0.89) and from good to excellent in sample 2 (function = 0.92, overall impact = 0.89, and severity of symptoms = 0.91).

Cronbach's alpha is considered a misestimator of reliability except in the unusual instance when all elements of a multiple-item measure are tau equivalent and all measurement error is random (18). Therefore, we also estimated reliability by means of coefficient H (19), which provides an estimate of the reliability of the construct when it is modeled with a structural equation model. The coefficient H value was 0.93.

Construct validity. Data from samples 1 and 2 were pooled (n = 292) to increase the power of the statistical analysis. Results from the *t*-test showed that the FIQR total score (*t*-test = 2.38, 89.8 df, *P* = 0.02) as well as the physical function (*t*-test = 2.71, 108 df, *P* = 0.01) and severity of symptoms (*t*-test = 2.45, 97.6 df, *P* = 0.02) domains distinguished between patients who were working outside the home and those on sick leave at the time of the study assessments (Figure 2). In contrast, the overall impact dimension did not differ significantly between the 2 groups (*t*-test <1, 109.5 df, *P* = 0.52).

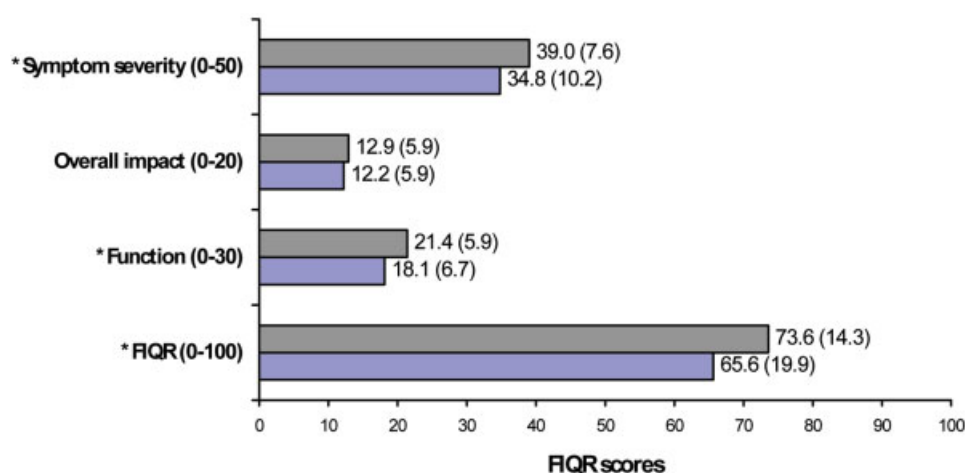


Figure 2. Construct validity: Revised Fibromyalgia Impact Questionnaire (FIQR) scores by patient report of employment status. Blue bars represent the group of patients working outside the home and gray bars represent the group of patients on sick leave; SDs are shown in parentheses. * = the *t*-test showed statistically significant differences between groups (*P* < 0.05).

DISCUSSION

This is the first study examining the dimensionality of the FIQR, currently considered the gold standard measure of functional impairment in FM (4,20). Although the initial PCA yielded several factors, a 1-factor solution appeared to capture more clearly the essence of the FIQR. Subsequently, we tested the statistical fit of a range of 5 viable factor models for the instrument in an independent FM sample. Because of the absence of previous analyses in the literature, our point of reference was the 1-factor solution found in the PCA.

Taken together, the 5 models failed to pass the strict chi-square test of model fit, indicating that the models did not capture all of the systematic variation in the data. Overall, the other fit indices (CFI, TLI, and RMSEA) provided strong and consistent evidence that covariance among the 21 FIQR items is best explained by a single underlying construct (functional impairment) with correlated residuals among items situated in the same instrument domain (models 2 and 3), which suggests that the FIQR items are complex in nature. In fact, the multidimensionality of the items was corroborated in the CBFA because, with the exception of 5 items in the specific factor of physical impairment, the other items had significant relationships with both the general and the specific factors.

To summarize, although the CFAs provided compelling evidence that a single, substantively meaningful construct underlies the responses to the 21 FIQR items, according to the CBFA it seems very reasonable to consider that the FIQR has a hierarchical structure in which all items are modeled as loading not only on a general factor of functional impairment, but also along with one of 3 specific factors corresponding to function, overall impact, and symptom severity. Therefore, it seems reasonable to compute and report not only an FIQR total score, but also 3 subscale scores, due to the multidimensional nature of the items. The present study demonstrates for the first time that the 5 items tap only into the underlying general functional impairment component and not the specific component of function that the FIQR was designed to measure. This finding suggests that a refinement of this domain might be required.

Concerning FIQR reliability, previous studies (5,8,9) have estimated internal consistency for the FIQR total score, but not for the subscale scores. When we calculated Cronbach's alpha for each subscale, good to excellent alpha values in the 3 FIQR domains were obtained. Internal consistency for the FIQR total score in sample 2 ($\alpha = 0.95$) was identical to that reported by Bennett and colleagues (5) and slightly higher than values found in subsequent studies by Ediz et al (8) and Srifi et al (9). At this point, taking the bifactor model into account, it is important to highlight that the summed total composite FIQR score is contaminated by the variance explained by the 3 specific factors. In such cases, according to Hancock and Mueller (19), it is more appropriate to estimate the reliability of the pure global latent factor because the variance attributed to the specific factors is controlled. As such, we estimated the reliability of the global latent factor via the coefficient H, which is considered a suitable alternative to

the conventional Cronbach's alpha. The obtained value of 0.93 is indicative of excellent reliability.

The FIQR total score and the physical impairment and severity of symptoms subscales were able to distinguish between patients on sick leave and those working at the time of the interview. The former reported greater functional impairment, physical deterioration, and symptom severity than the latter. This finding adds strong support to the discriminant validity of the FIQR. Bennett et al (5) had previously shown that FM patients scored significantly higher on the FIQR than healthy individuals, patients with either rheumatoid arthritis or systemic lupus erythematosus, and patients who have major depression. The absence of significant differences in the overall impact domain was an unexpected result. The simple conclusion is that this domain is less sensitive in detecting differences between specific subgroups of FM patients. Future studies should focus more thoroughly on the construct validity of the 3 FIQR domains.

There is some divergence among the FIQR psychometric studies. For example, the Spanish and Moroccan FM patients were severely impaired according to FIQ severity categorization (21), which considers a total FIQ score ≥ 59 as indicative of severe impairment. The Spanish patients recruited for our study were approximately 10 points above the mean FIQR scores for 202 North American FM patients (56.6) and 87 Turkish FM patients (55.2 at the first visit and 57.16 at the second visit), with both situated in the moderate impairment category (FIQ total score ≥ 39 to < 59). It is also noticeable that the FIQR items in Spain seem to contribute to a lesser degree to the homogeneity (22) of the instrument than in the US, considering we found corrected item-total correlations that ranged from 0.45–0.74 in sample 1 and from 0.52–0.77 in sample 2, whereas Bennett et al (5) reported values that oscillated between 0.56 and 0.93.

Because of the relatively small sample size, conclusions based on the present data must be considered preliminary until more factor analytical studies appear in the literature. Some important psychometric aspects of the FIQR remain unknown and should be examined in the future. First, whether FIQR dimensionality is invariant in male and female FM patients and in patients from distinct cultures should be analyzed in large, multinational samples. Our findings are limited to Spanish FM patients, and the dimensionality of the instrument might vary as a function of national origin, region, or language. Second, by means of structural equation modeling, researchers would be able to estimate the relationships between the 3 specific factors and external criteria. If the global and specific FIQR factors were associated with other constructs in a different manner, this would suggest that reporting a total FIQR score results in biased relationships with external criteria. Third, because of the cross-sectional design of our study, we could not examine the responsiveness, the smallest detectable change, or the minimum clinically important difference for scoring the FIQR, as has been done for the FIQ (21). Fourth, to date, FIQR psychometric analyses have used classic test theory as a framework. However, this framework offers no means to gauge the quality of individual FIQR items and response options across differ-

ent levels of the trait. The use of methods based on item response theory (23) or structural equation modeling might provide detailed information about the functioning of each FIQR item and would allow assessment of differential item functioning.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Luciano had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Luciano.

Acquisition of data. Calandre, Rodriguez-Lopez.

Analysis and interpretation of data. Luciano, Aguado, Serrano-Blanco.

REFERENCES

- Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: report of the Multicenter Criteria Committee. *Arthritis Rheum* 1990;33:160–72.
- Wolfe F, Clauw DJ, FitzCharles MA, Goldenberg DL, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. *Arthritis Care Res (Hoboken)* 2010;62:600–10.
- Mease PJ, Arnold LM, Choy EH, Clauw DJ, Crofford LJ, Glass JM, et al. Fibromyalgia syndrome module at OMERACT 9: domain construct. *J Rheumatol* 2009;36:2318–29.
- Boomershine CS. A comprehensive evaluation of standardized assessment tools in the diagnosis of fibromyalgia and in the assessment of fibromyalgia severity. *Pain Res Treat* 2012;2012:653714.
- Bennett RM, Friend R, Jones KD, Ward R, Han BK, Ross RL. The Revised Fibromyalgia Impact Questionnaire (FIQR): validation and psychometric properties. *Arthritis Res Ther* 2009;11:R120.
- Burckhardt CS, Clark SR, Bennett RM. The Fibromyalgia Impact Questionnaire: development and validation. *J Rheumatol* 1991;18:728–33.
- Bennett RM. The Fibromyalgia Impact Questionnaire (FIQ): a review of its development, current version, operating characteristics and uses. *Clin Exp Rheumatol* 2005;23 Suppl 39: S154–62.
- Ediz L, Hiz O, Toprak M, Tekeoglu I, Ercan S. The validity and reliability of the Turkish version of the Revised Fibromyalgia Impact Questionnaire. *Clin Rheumatol* 2011;30:339–46.
- Srifi N, Bahiri R, Rostom S, Bendeddouche I, Lazrek N, Hajjaj-Hassouni N. The validity and reliability of the Moroccan version of the Revised Fibromyalgia Impact Questionnaire. *Rheumatol Int* 2013;33:179–83.
- Calandre EP, Vilchez JS, Molina-Barea R, Tovar MI, Garcia-Leiva JM, Hidalgo J, et al. Suicide attempts and risk of suicide in patients with fibromyalgia: a survey in Spanish patients. *Rheumatology (Oxford)* 2011;50:1889–993.
- Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw-Hill; 1994.
- Kaiser HF. Analysis of factorial simplicity. *Psychometrika* 1974;39:31–6.
- Bartlett MS. A note on the multiplying factors for various chi-square approximations. *J R Stat Soc Series B Stat Methodol* 1954;16:296–8.
- Fabrigar LR, Wegener DT. *Understanding statistics: exploratory factor analysis*. New York: Oxford University Press; 2012.
- Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equat Model* 1999;6:1–55.
- Kline P. *A handbook of test construction*. London: Methuen; 1986.
- Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res* 2007;16:19–31.
- Sijsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 2009;74:107–9.
- Hancock GR, Mueller RO. Rethinking construct reliability within latent variable systems. In: Cudeck R, Du Troit S, Sorbom D, editors. *Structural equation modeling: present and future*. Lincolnwood (IL): Scientific Software International; 2001. p. 195–216.
- Williams DA, Arnold LM. Measures of fibromyalgia: Fibromyalgia Impact Questionnaire (FIQ), Brief Pain Inventory (BPI), Multidimensional Fatigue Inventory (MFI-20), Medical Outcomes Study (MOS) Sleep Scale, and Multiple Ability Self-Report Questionnaire (MASQ). *Arthritis Care Res (Hoboken)* 2011;63 Suppl 11:S86–97.
- Bennett RM, Bushmakina AG, Cappelleri JC, Zlateva G, Sadosky AB. Minimal clinically important difference in the Fibromyalgia Impact Questionnaire. *J Rheumatol* 2009;36:1304–11.
- Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 2nd ed. Oxford: Oxford University Press; 1995.
- Thomas ML. The value of item response theory in clinical assessment: a review. *Assessment* 2011;18:291–307.