Measurement invariance of the Illness Invalidation Inventory (3*I) across language, rheumatic disease, and gender

Corresponding author: Marianne Kool, PhD, PO Box 80140, 3508 TC Utrecht, The Netherlands. E-mail: m.b.kool@uu.nl; Tel.: +31 30 253 2387; Fax: +31 30 253 4718

Marianne B Kool,[1,10] Rens van de Schoot, [2] Isabel López-Chicheri García, [3] Ricarda Mewes, [4] José AP Da Silva, [5] Karoline Vangronsveld, [6] Andreas AJ Wismeijer,[7] Mark A Lumley, [8] Henriët van Middendorp,[9] Johannes WJ Bijlsma,[10] Geert Crombez,[6] Winfried Rief,[4] Rinie Geenen [1,10]

[1]Department of Clinical and Health Psychology, Utrecht University, Utrecht, Netherlands, [2]Department of Methods and Statistics, Utrecht University, Utrecht, Netherlands, [3]Department of Personality, Evaluation and Psychological Treatment, University of Granada, Granada, Spain, [4]Department of Clinical Psychology and Psychotherapy, Philipps University Marburg, Marburg, Germany, [5]Department Medical Service III and Rheumatology, University Hospital Coimbra, Coimbra, Portugal, [6]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium, [7]Department of Developmental and Clinical Psychology, Tilburg University, Tilburg, Netherlands, [8]Department of Psychology, Wayne State University, Detroit, USA, [9]Department of Medical Psychology, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands, [10]Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht, Netherlands.

Key words: Psychology, Fibromyalgia/Pain Syndromes, Osteoarthritis, Rheumatoid arthritis,

Word count: 2973 (Introduction - Discussion); Abstract 244, 22 references, 5 tables, 1 figure, 1 supplementary file

ABSTRACT

Objectives: The Illness Invalidation Inventory (3*I) assesses patients' perception of responses of others that are perceived as denying, lecturing, not supporting, and not acknowledging the condition of the patient. It includes two factors: 'discounting' and 'lack of understanding'. In order to use the 3*I to compare and to pool scores across groups and countries, the questionnaire must have measurement invariance; that is, it should measure identical concepts with the same factor structure across groups. The aim of this study was to examine measurement invariance of the 3*I across rheumatic diseases, gender, and languages.

Methods: Participants with a rheumatic disease from various countries completed an on-line study, which included the 3*I and which was presented in Dutch, English, French, German, Portuguese, and Spanish; 6,057 people with rheumatic diseases participated. Single and multiple group confirmatory factor analyses were used to test the factorial structure and measurement invariance of the 3*I with M*plus*.

Results: The model with strong measurement invariance, i.e. equal factor loadings and thresholds (distribution cut points) across gender and rheumatic disease (fibromyalgia versus other rheumatic diseases) had the best fit estimates for the Dutch version, and had good fit estimates across the six language versions.

Conclusions: The 3*I showed measurement invariance across gender, rheumatic disease, and language. Therefore, it is appropriate to compare and to pool scores of the 3*I across groups. Future research may use the questionnaire to examine antecedents and consequences of invalidation as well as the effect of treatments targeting invalidation.

INTRODUCTION

Invalidation, defined as the perception of cognitive, affective, and behavioural responses of others that are judged to be denying, lecturing, not supporting, and not acknowledging the condition of the patient,[1] is problematic for some patients with rheumatic diseases. Symptoms of rheumatic diseases such as pain, fatigue, and stiffness are mostly invisible and because of this, people in the social environment of the patient might forget or misjudge the burden and consequences of the illness.[2] When there is no clinical or laboratory evidence to account for the symptoms of the rheumatic illness, such as in fibromyalgia,[3] this can provoke even more serious disbelief and distrust towards the patients.[4] Indeed, patients with fibromyalgia reported that invalidation is a major issue in their lives, adding a burden to the symptoms.[1, 5] Moreover, invalidation could increase the risk of becoming more physically impaired and depressed,[6] thus highlighting the need for attending to invalidation in research and clinical settings.[7]

The Illness Invalidation Inventory (3*I) is a self-report questionnaire that assesses patients' invalidation. An initial evaluation suggested internal consistency and concurrent validity of the inventory.[6] To be able to use the 3*I in epidemiological studies, to make comparisons of invalidation across patient groups and countries, and to examine its antecedents and consequences, the questionnaire must measure identical constructs with the same structure across groups. This means that the factor structure, factor loadings, and thresholds should be comparable across groups,[9] which is called measurement invariance. Thresholds are the points on the unobserved normal distribution where, on average, respondents vary between two different response options.[8] When measurement invariance is absent, groups or subjects respond differently to items, and factor means cannot validly be compared across groups. As yet, it is unclear whether the 3*I shows measurement invariance across rheumatic diseases, gender, and languages.

The 3*I assesses invalidation by the spouse, family, medical professionals, work environment, and social services using eight items for each of these five social sources (Figure 1). The structure for each social source comprises two factors; discounting (5 items) and lack of understanding (3 items). Discounting represents active negative social responses including disbelieving, admonishing, dismissing inability to work, not acknowledging symptom fluctuations, and offering unusable advice. Lack of understanding reflects a lack of positive social responses such as not recognising, comprehending, and emotionally supporting the patient or illness. To date, the Dutch version of the 3*I has been validated in patients with rheumatoid arthritis and fibromyalgia,[6] but measurement invariance across rheumatic diseases, gender, and language versions of the 3*I has not been tested. Furthermore, qualitative studies suggest that invalidating experiences of patients with rheumatic diseases are common in many countries.[4, 5, 10, 11] However, the study of invalidation across countries is only legitimately possible after the equivalence of the language versions (other than Dutch) of the 3*I has been established.

The first aim of the current study was to test the factor structure in a large sample of Dutch patients with several rheumatic diseases and to examine measurement invariance between patients with fibromyalgia and patients with other rheumatic diseases and between men and women. The second aim was to examine measurement invariance between different language versions (Dutch, English, French, German, Portuguese, and Spanish).

METHODS

Participants

Participants were 6,057 people with rheumatic diseases from different countries. Inclusion criteria were (1) the self-report of a rheumatic disease, (2) the report that the disease was

diagnosed by a medical specialist, general practitioner, or nurse, (3) being 18 years or older, and (4) speaking Dutch, English, French, German, Portuguese, or Spanish. Table 1 shows the characteristics of the participants, who had a mean age of 45.7 years and were mostly female (88%). With the exception of the German respondents, patients with fibromyalgia constituted the largest percentage of respondents (40-76% across different languages). The German version of the questionnaire was completed by patients with a wide range of conditions, particularly systemic lupus erythematosus (29%) and ankylosing spondylitis (25%). Overall, patients with osteoarthritis (8-28%), rheumatoid arthritis (6-20%), and Sjögren's syndrome (4-21%) were well represented in the study.

Procedure

The study was conducted according to the principles of the Declaration of Helsinki [12] and was tested and approved by the Medical Ethical Committee of the University Medical Center Utrecht.

The software program "Netquestionnaires" [13] was used to develop online international questionnaires. An online version was first developed for the Dutch version, which was pretested among a small sample of Dutch patients. After translation, the other language versions were developed uniformly and pre-tested by a small expert group.

Participants were invited via a recruitment notice on websites of patient associations for rheumatic diseases. Patient associations in Dutch, English, French, German, Portuguese, and Spanish language countries were asked to put the recruitment notice on their website. The text of this notice was similar across countries and patient associations. It is unknown whether (some) patient organizations took additional actions to bring the call to the notice of patients. On the website, participants could choose one of the six language versions of the study. The recruitment notice included information about the aim and content of the study, inclusion

criteria, duration of participation (about 20 minutes), and a hyperlink to the online questionnaire. Participants could decide to participate after being informed about the study, and were able to stop at any point if they desired. Of 8,293 participants who began filling out the questionnaire, 6,057 (77%) completed it and provided the data analyzed here. About 80% of drop outs were participants who stopped filling out the questionnaire while responding to the first demographic questions.

Table 1 Demographic characteristics of patients with rheumatic diseases for the six language versions

Characteristics	**Dutch**	**English**	**French**	**German**	**Portuguese**	**Spanish**
Sample size	1855	774	735	513	727	1453
Gender; female, n (%)	1608 (86%)	717 (93%)	688 (94%)	403 (79%)	628 (86%)	1308 (90%)
Age (yrs), mean (SD)	46.9 (12.7)	47.2 (11.3)	46.5 (11.4)	46.5 (12.3)	42.9 (12.6)	44.1 (11.4)
Rheumatic disease, *n* (%)						
Ankylosing spondylitis	236 (13%)	6 (1%)	52 (7%)	129 (25%)	61 (8%)	104 (7%)
Bursitis / Tendinitis	89 (5%)	93 (12%)	45 (6%)	22 (4%)	85 (12%)	148 (10%)
Fibromyalgia	737 (40%)	591 (76%)	482 (65%)	65 (13%)	303 (42%)	778 (54%)
Gout or Pseudogout	31 (2%)	7 (1%)	1 (0.1%)	6 (1%)	9 (1%)	2 (0.3%)
Juvenile arthritis	6 (0.3%)	6 (1%)	1 (0.1%)	8 (2%)	15 (2%)	49 (3%)
Osteoarthritis	518 (28%)	155 (20%)	176 (24%)	40 (8%)	105 (14%)	212 (15%)
Polymyalgia rheumatica	17 (1%)	4 (1%)	15 (2%)	4 (1%)	5 (1%)	37 (3%)
Psoriatic arthritis	102 (6%)	24 (3%)	10 (1%)	26 (5%)	31 (4%)	34 (2%)
Raynaud’s phenomenon	76 (4%)	70 (9%)	78 (11%)	79 (15%)	24 (3%)	71 (5%)
Rheumatoid arthritis	323 (17%)	137 (18%)	46 (6%)	68 (13%)	148 (20%)	271 (19%)

Sarcoidosis	3 (0.2%)	2 (0.3%)	2 (0.3%)	9 (2%)	2 (0.3%)	1 (0.1%)
Scleroderma	26 (1%)	10 (1%)	4 (1%)	45 (9%)	9 (1%)	27 (2%)
Sjögren's syndrome	122 (7%)	34 (4%)	91 (12%)	106 (21%)	25 (3%)	137 (9%)
Systemic lupus erythematosus	197 (6%)	18 (2%)	125 (17%)	146 (29%)	122 (17%)	233 (16%)
Other rheumatic disease	244 (13%)	70 (9%)	65 (9%)	72 (14%)	87 (12%)	198 (14%)

Note. Percentages of rheumatic diseases can exceed 100% because participants may have more than one rheumatic disease

Instruments

The on-line study included items about demographic characteristics and the Illness Invalidation Inventory.

The Illness Invalidation Inventory (3*I) [6] measures invalidation by each of 5 sources (spouse, family, medical professionals, work environment, and social services). It consists of 2 factors: discounting (5 items) and lack of understanding (3 items; see Figure 1). Participants indicate on a 5-point scale (1=never, 2=seldom, 3=sometimes, 4=often, 5=very often) how frequently during the past year people within each category responded to them in the described way. A source category that does not apply can be skipped. An initial validation study in Dutch patients demonstrated good reliability and validity of the 3*I.[6]

Translation of the 3*I was done by the forward-and-backward translation method. First, the Dutch version of the 3*I was translated by a native English and Dutch speaker to English. When consensus was reached about the English version, another native English and Dutch speaker translated the English version back to Dutch. After consensus was reached between translators and researchers, the final English version of the questionnaire was determined. The English version of the questionnaire was used for the translation to other languages. For each translation, the forward-and-backward translation method was used by two native English speakers and two native speakers of the other language.

Measurement invariance

The first step to determine measurement invariance is to specify the model (adequate structure) of the instrument.[14] With confirmatory factor analyses (CFA) the factor structure of the model is tested for each group. The second step is to check whether the best fitting factor model is adequate and equal across groups. This can be tested first, by comparing the numerical values of the factor loadings across groups, which should be similar (weak

invariance); and, second, by comparing whether the thresholds are similar across groups (scalar invariance).[15] For straightforward interpretation of latent variable means and patterns of correlations across groups, both the factor loadings and thresholds should be similar across groups (strong invariance).[16]

When weak invariance is not supported, this could mean that one or more of the common factors have different meanings across the population groups [17] or that a subset of the factor loading estimates for a group is biased due to extreme response style. When strong invariance is not supported, differential additive response bias or differential acquiescence response styles might be the problem.[18]

Statistical analysis

Single and multiple group CFAs were used to test the factorial structure and measurement invariance by means of M*plus* 6.11.[19]

First, the Dutch data were used to check whether the model displayed in Figure 1 with two factors fits to the data better than a one-factor model. Furthermore, we investigated whether the model fits better to the data with continuous or with ordered categorical indicators. Because a model including all sources of the 3*I with categorical indicators was too complex for M*plus* to compute, the model was run for each source separately (spouse, family, medical professionals, work environment, and social services). In the Results section, the statistics for the source 'family' are presented because this source applies to almost all participants. Results for the other four sources are presented in a Supplementary file.

Due to non-normally distributed item scores, a robust weighted least squares estimator (WLSMV) was used. Full information maximum likelihood estimation was used to include participants with a score on at least one item of a subscale.[20] To assess model fit, we used the comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of

approximation (RMSEA). Cut-off values for fit were considered adequate if CFI and TLI values are > 0.90 and RMSEA < 0.08. The Bayesian information criterion (BIC) was used to compare competing models. A lower BIC indicates a better trade-off between model fit and model complexity. Because the BIC value is not estimated by the WLSMV estimator, all models were repeated using a maximum likelihood estimator (MLR), again using a correction for non-normality.

In a second step of analyses, measurement invariance across rheumatic diseases was tested in the Dutch data for patients with only fibromyalgia versus patients with one other rheumatic disease (n = 890). The other 548 patients who reported having two or more co-morbid rheumatic diseases were excluded from this analysis. Third, measurement invariance across gender was tested in the Dutch data. Fourth, measurement invariance was tested across languages in the international data using a similar procedure.

In each test of measurement invariance, three models were analyzed; Model 1 includes constrained factor loadings, thresholds free (weak invariance), Model 2 includes constrained thresholds, factor loadings free (scalar invariance), and Model 3 includes both constrained factor loadings and thresholds (strong invariance) [21].

In a last step of analysis, standardized factor loadings, threshold values, and internal consistency (Cronbach's alpha) were examined.

RESULTS

Factor structure of the Illness Invalidation Inventory

For several items of the 3*I, the score distribution across the response categories were skewed. Therefore, one- and two-factor models with continuous factor indicators were compared to a one- and two-factor model where the items were defined as being categorical. Table 2 shows outcomes of the four CFA for the 3*I source 'family' factor structure in the

Dutch data. The two-factor model with categorical items provided the better trade-off between model fit and model complexity according to the BIC criterion and it had adequate fit estimates. Analyses for the four other social sources were repeated and the two-factor model with categorical items also showed the best fit for each source (Supplementary file, Table S1). This model has been used in subsequent analyses.

Table 2 Fit statistics of confirmatory factor analyses of the 3*I source 'family' for the one and two-factor solution including continuous or categorical factor indicators in the Dutch sample (n = 1855)

Model:	χ^2	df	CFI	TLI	RMSEA	BIC
One-factor continuous	1040.94	20	0.90	0.86	0.17	37236
One-factor categorical	16064.02	390363	0.97	0.95	0.20	34158
Two-factor continuous	279.29	19	0.97	0.96	0.09	36482
Two-factor categorical	15931.29	390397	0.99	0.99	0.10	33615

Note. CFI = comparative fit index, TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation, BIC = Bayesian information criterion. The df-value differs between the continuous and categorical models due to a high number of parameter estimators in the categorical model, since there are k (response categories)-1 thresholds

Measurement invariance 3*I across rheumatic diseases

Table 3 shows outcomes of the CFA for the source 'family' of the 3*I for fibromyalgia versus other rheumatic diseases in the Dutch data. The three models were tested for each source of

the 3*I. Although Model 1a had the lowest χ^2-value and RMSEA, and the largest CFI and TLI, the fit indices were also acceptable for Model 3a. The BIC value was lowest for Model 3a, which shows that Model 3a with both the factor loadings and thresholds constrained is simpler than Model 1a; it is the preferred model because it has a better trade-off between model fit and model complexity. The BIC value was also lowest for Model 3a among the other sources (Supplementary file, Table S2). Therefore, Model 3a was chosen as the best model indicating measurement invariance across rheumatic diseases.

Table 3 Test of measurement invariance of source 'family' of the 3*I of rheumatic disease (fibromyalgia versus other rheumatic disease) and gender in the Dutch sample, and of language in the international sample

Source 3*I	χ^2	**df**	**CFI**	**TLI**	**RMSEA**	**BIC**
Rheumatic disease (n = 1314)						
Model 1a: thresholds free	16027.22	780967	0.99	0.99	0.08	25352
Model 2a: factor loadings free	16928.29	780997	0.98	0.98	0.11	25319
Model 3a: factor loadings + thresholds fixed	16430.65	781001	0.98	0.99	0.10	25281
Gender (n = 1855)						
Model 1b: thresholds free	18223.51	780943	0.99	0.99	0.09	35368

Model 2b: factor loadings free	18325.51	780966	1.00	1.00	0.06	35236
Model 3b: factor loadings + thresholds fixed	18427.46	780975	1.00	1.00	0.05	35185
Language (*n* = 6027)						
Model 1c: thresholds free	85824.25	2342233	0.99	0.99	0.08	139064
Model 2c: factor loadings free	88422.58	2342289	0.98	0.99	0.10	140846
Model 3c: factor loadings + thresholds fixed	89713.99	2342336	0.98	0.99	0.09	140984

Note. CFI = comparative fit index, TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation, BIC = Bayesian information criterion.

Measurement invariance 3*I across gender

Measurement invariance was also tested across gender in the Dutch data. Table 3 shows the fit-estimates for the source 'family'. Model 3b (strong invariance) best fitted the data: it had adequate CFI, TLI and RMSEA values and the lowest BIC value. Model 3b showed also the lowest BIC value among the other sources (Supplementary file, Table S3). This supports measurement invariance across gender.

Measurement invariance 3*I across languages

In the separate English, French, German, Portuguese, and Spanish data sets, CFA was conducted to examine the two-factor structure of the 3*I for each language version of the 3*I as compared to the Dutch factor structure. The two-factor model showed adequate fit estimates for each language version of the 3*I (results are not shown). Because this result confirmed the equivalence of factor structures of the 3*I across languages, it was feasible to study measurement invariance of the 3*I across the languages.

A multiple group model including all data sets was created to compare Models 1c, 2c and 3c across languages. Fit estimates are shown in Table 3 for the source 'family' of the 3*I. All Models (1c, 2c, and 3c) showed adequate CFI and TLI estimates. The fit-estimates showed the best fit for model 1c (weak invariance): in general it had the lowest χ^2-*value*, RMSEA, and BIC and the largest CFI and TLI. This model 1c (constrained factor loading, thresholds free) indicates that the factor loadings are invariant across languages, but that threshold values are non-invariant across languages.

To illustrate this finding, consider the standard factor loadings obtained with Model 1c (factor loadings free, Table 4). The factor loadings differ barely between language versions of the 3*I. Thus, constraining these factor loadings to be equal resulted in a simpler model (i.e., less parameters have to be estimated) and it yields a better fit (Table 3).

Table 5 shows the thresholds of items 1 and 2 for the source 'family' of the 3*I in each language version. The results indicate differences between languages, for example item 2 threshold 2 shows -0.25 for the French version and 0.40 for the German version. If we constrain the thresholds to be equal (Model 3c), the BIC indicates that, although model 3C is simpler, the fit has worsened.

Table 4 Standardized factor-loadings for 3*I for each language version

Items per factor	Factor loadings					
	Dutch	**English**	**French**	**German**	**Portuguese**	**Spanish**
Discounting by family						
Item 1	0.74	0.77	0.74	0.84	0.76	0.76
Item 2	0.83	0.84	0.87	0.91	0.84	0.91
Item 4	0.48	0.50	0.38	0.71	0.56	0.46
Item 6	0.89	0.93	0.92	0.93	0.92	0.95
Item 7	0.86	0.93	0.91	0.92	0.91	0.93
Lack of understanding by family						
Item 3	0.85	0.87	0.90	0.91	0.89	0.93
Item 5	0.80	0.86	0.88	0.89	0.87	0.91
Item 8	0.77	0.80	0.86	0.84	0.82	0.85

Note. Cross-loadings are set to zero in confirmatory factor analyses.

Table 5 Thresholds item 1 and 2 of the source 'family' of the 3*I of each language version

Items per factor	**Dutch**	**English**	**French**	**German**	**Portuguese**	**Spanish**
Family discounting						
Item 1 threshold 1	-0.78	-1.07	-0.79	-0.50	-0.32	-0.68
Item 1 threshold 2	-0.19	-0.57	-0.36	0.04	0.13	-0.26
Item 1 threshold 3	0.73	0.35	0.35	0.82	0.95	0.33
Item 1 threshold 4	1.66	1.13	1.13	1.63	1.68	0.85
Item 2 threshold 1	-0.47	-0.68	-0.70	-0.18	-0.50	-0.51
Item 2 threshold 2	0.20	-0.15	-0.25	0.40	-0.01	-0.17

Item 2 threshold 3	0.94	0.55	0.30	1.01	0.64	0.31
Item 2 threshold 4	1.73	1.29	1.04	1.72	1.37	0.78

Note. A 5-point response scale contains four thresholds, because thresholds divide the distribution into distinct categories equal to the number of categories minus one.

The fit of Model 3c could be improved through partial measurement invariance (setting some thresholds "free", but constraining others). To identify non-invariant thresholds, both the Modification Index (MI) for a parameter, which gives the expected drop in the model's Chi-square value if the parameter is freely estimated, and the value of thresholds across languages can be studied. Ideally, one or two items, or one or two languages would be identified explaining non-invariant thresholds. In that case, a solution would be to delete these items, or to deal differently with computation of factors in this language. In our study, the German data seem most different. However, close inspection of the MI and the (5 times 32) threshold values across languages did not show significant MI's or specific threshold patterns. Therefore, determining the degree of partial measurement invariance was not attainable. Because the model assuming strong measurement invariance still has a good fit to the data (i.e., most model fit indices are well above the cut-off values) and differences in threshold values probably cancel each other out between the different languages[8, 22], we concluded that Model 3c applies for all social sources (Supplementary file, Table S4). The Cronbach's alpha per language indicated that the internal consistency of the two 3*I factors is good; between .76 and .95 for the factor 'discounting' and between .78 and .93 for the factor 'lack of understanding'. This provides additional support for using Model 3c.

DISCUSSION

This study investigated the measurement invariance of the 3*I across rheumatic disease,

gender, and language version. Our results confirm the validity of a two-factor structure of the 3*I comprising discounting and lack of understanding for each source (spouse, family, medical professionals, work environment, and social services) in patients with rheumatic diseases. Strong measurement invariance (i.e., constrained factor loadings and thresholds) across rheumatic disease (fibromyalgia versus other rheumatic disease) and across gender was established. Strong measurement invariance across the six language versions of the 3*I was also supported by adequate fit estimates and by good internal consistency of the factors for each language. In most cases, the BIC supported strong measurement invariance, but not all estimates showed the best fit for this model. Because most models under investigation had a moderate to high RMSEA (.08-.14) and about equal CFI and TLI fit statistics, our model selection was largely based on BIC. Although the fit of the model with constrained factor loadings and thresholds was not the best across language, measurement invariance of the 3*I for languages was not rejected. Because all models (weak, scalar, and strong invariance) showed good fit estimates, it is acceptable to conclude that the different language versions of the 3*I are comparable.

This study has some limitations. First, diagnoses were self-reported by patients, and there was no certification by a medical specialist. Second, participants were recruited through the internet, which may have led to a younger sample from a higher social economic class. However, this is less of a problem for the examination of measurement invariance, because the instrument should be invariant whether patients do or do not have the disease, are young or old, etc. Third, no expert committee was involved in the translation procedure of the 3*I.

The 3*I is the first instrument to quantify invalidation in patients with rheumatic diseases. Results of our study indicate that comparisons between diseases, gender and language version of the 3*I are possible, and are likely to be not affected by different response styles or different interpretations of items. This is an advantage when examining

antecedents and consequences of invalidation as well as the effects of treatments targeting invalidation. In clinical practice, the questionnaire can be used to assess invalidation in individual patients, which will help therapists to understand and treat patients' problems.

Acknowledgement We would like to thank all participants for their contribution to this study, the patient associations for help in recruiting, and Joop Hox for his expertise and suggestions.

Funding This study was funded by a grant from the Dutch Arthritis Association (Reumafonds).

Competing Interest: None declared

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Kool MB, Van Middendorp H, Boeije H, et al. Understanding the lack of understanding. Invalidation from the perspective of the patient with fibromyalgia. *Arthritis Rheum-Arthritis Care Res* 2009;**61**:1650-6.

2. Cunningham MM, Jillings C. Individuals' descriptions of living with fibromyalgia. *Clin Nurs Res* 2006;**15**:258-73.

3. Wolfe F, Ross K, Anderson J, et al. The prevalence and characteristics of fibromyalgia in the general population. *Arthritis Rheum* 1995;**38**:19-28.

4. Asbring P, Narvanen AL. Women's experiences of stigma in relation to chronic fatigue syndrome and fibromyalgia. *Qual Health Res* 2002;**12**:148-60.

5. Zangi HA, Hauge M, Steen E, et al. "I am not only a disease, I am so much more". Patients with rheumatic diseases' experiences of an emotion-focused group intervention. *Patient Educ Couns* 2011;**85**:419-24.

6. Kool MB, Van Middendorp H, Lumley MA, et al. Lack of understanding in fibromyalgia and rheumatoid arthritis: The Illness Invalidation Inventory (3*I). *Ann Rheum Dis* 2010;**69**:1990-5.

7. Geenen R. Educating patients and the struggle for validation of fibromyalgia syndrome: The Dutch way. *J Musculoskelet Pain* 2009;**17**:80-5.

8. Sass DA. Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *J Psychoeduc Assess* 2011;**29**:347-63.

9. Mewes R, Christ O, Rief W. Comparing assessment results between cultures. Submitted for publication, 2011.

10. Arnold LM, Crofford LJ, Mease PJ, et al. Patient perspectives on the impact of fibromyalgia. *Patient Educ Couns* 2008;**73**:114-20.

11. Werner A, Malterud K. It is hard work behaving as a credible patient: Encounters between women with chronic pain and their doctors. *Soc Sci Med* 2003;**57**:1409-19.

12. World Medical Association (WMA). Declaration of Helsinki. Seoul: WMA; 2008. Available from: http://www.wma.net/e/policy/b3.htm.

13. NetQuestionnaires. Manual NETQ Internet Survey 6.0. Utrecht: NetQuestionnaires Nederland BV, 2007.

14. Dimitrov DM. Testing for factorial invariance in the context of construct validation. *Meas Eval Counsel Dev* 2010;**43**:121-49.

15. Reeve BB, Fayers P. Applying item response theory modeling for evaluating questionnaire item and scale properties. In: Fayers P, Hays RD, editors. Assessing Quality of Life in Clinical Trials: Methods of Practice. 2nd Edition ed. Oxford: Oxford University Press; 2005: 55-73.

16. Bowden SC, Weiss LG, Holdnack JA, et al. Equivalence of a measurement model of cognitive abilities in US standardization and Australian neuroscience samples. *Assessment* 2008;**15**:132-44.

17. Gregorich SE. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Med Care* 2006;**44**:S78-94.

18. Byrne B, Watkins D. The issue of measurement invariance revisited. *J Cross Cult Psychol* 2003;**34**:155-75.

19. Muthén LK, Muthén BO. Mplus User's Guide, Sixth edition. Los Angeles, CA: Muthén & Muthén 2010.

20. Enders CK, Bandalos DL. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Struct Equ Modeling* 2001;**8**:430-57.

21. Van de Schoot R, Lugtig P, Hox J. A checklist for testing measurement invariance. Eur J Dev Psychol 2012;9:486-92.

22. Byrne B, Shavelson R, Muthén B. Testing for the equivalence of factor covariance and mean structures - The issue of partial measurement invariance. *Psychol Bull* 1989;**105**:456-66.

Figure 1 Factor structure of the Illness invalidation Inventory (3*I)

Note. Items of the factor 'lack of understanding' have reversed response scores.