Journal of Clinical Epidemiology 174 (2024) 111488

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

The Flare-OA-16 questionnaire measuring flare in knee and hip osteoarthritis in the patient perspective: scale reduction and validation using a Rasch model

Fabiana Queiroga[a], Jonathan Epstein[a,b,*], Marie-Line Erpelding[a,b], Elisabeth Spitz[a], Jean-Francis Maillefert[c], Bruno Fautrel[d,e], Leigh F. Callahan[f], David J. Hunter[g], Francis Guillemin[a,b]

[a]*CIC Clinical epidemiology, CHRU Nancy, Inserm, Université de Lorraine, Nancy, France*
[b]*Inserm, CHRU Nancy, Université de Lorraine, Clinical Epidemiology, Nancy, France*
[c]*Department of Rheumatology, University Hospital, Dijon, France*
[d]*Sorbonne Université – Assistance Publique Hôpitaux de Paris, Rheumatology Dept, Pitié-Salpêtrière Hospital, Paris, France*
[e]*Inserm, Institut Pierre Louis d'Epidémiologie et Santé Publique, Paris, France*
[f]*Thurston Arthritis Research Center, University of North Carolina, Chapel Hill, NC, USA*
[g]*Rheumatology Department, Sydney Musculoskeletal Health, Kolling Institute, and Royal North Shore Hospital, University of Sydney, Sydney, Australia*

Accepted 25 July 2024; Published online 30 July 2024

Abstract

Objectives: The recent Flare-OA questionnaire measuring flare in knee and hip osteoarthritis (OA) (19 items in 5 domains, numerical rating scale) showed good psychometric properties along with classical test theory. This study aimed to determine its scaling properties by Rasch analysis and to present evidence for a refined scalable version.

Study Design and Setting: The participants were 398 subjects (mean age 64 years [standard deviation = 8.1], 70.4% women) recruited from Australia, France, and the United States, with clinically and radiologically symptomatic knee or hip OA, who completed an online survey. The sample was split into derivation and validation subsamples, stratified by country and joint. Rasch analysis examined differential item functioning (DIF) for sex, age, country, and joint. A confirmatory factor analysis and an analysis of convergent validity were performed to document the psychometric properties of the short version.

Results: To fit the Rasch model, we reordered thresholds of answering modalities when necessary. Two items were removed. A local dependency between 2 items was solved by combining items modalities into a super-item. A uniform DIF (expected and nonremoved) was identified for one item that was split by joint, and a nonuniform DIF for one item for age and country (removed). The person-item threshold distribution showed a well-focused scale; the confirmatory factor analysis and the analysis of convergent validity showed good fit indicators for the short version.

Conclusion: The Rasch analysis was helpful in guiding the decision to refine the measurement instrument. After analysis, the 16-item Flare-OA self-report questionnaire is available for use in clinical research.

Keywords: Flare; Oesteoarthritis; Knee and hip; Patient's perspective; Self-reporting questionnaire; Rasch model

1. Introduction

Osteoarthritis (OA) is the most common form of arthritis, involving chronic impairment, and structural alterations of joints. According to the Global Burden of Disease study, OA is responsible for 2.2% of total global years lived with disability and OA knee and hip contributed 60.9% and 5.5% of OA years lived with disability, respectively [1,2]. OA is typically characterized as a progressive disease marked by a slow, steady decline of function. There is increasing recognition that ''acute-on-chronic'' episodes and ''flare-ups'' of more severe pain are part of the disease process [3,4].

Literature reviews on knee and hip OA point out that there is substantial evidence about how common the disease is [5–8] and encourage improved measurement of aspects

Funding: This work was supported partly by the French PIA project ''Lorraine Université d'Excellence'', reference ANR-15-IDEX-04-LUE.

* Corresponding author. INSPIIRE, Campus Brabois Santé, Université de Lorraine, 9, avenue de la Forêt de Haye CS 50184, 54505 Vandoeuvre-lès-Nancy, France.
E-mail address: fabiana.queiroga@univ-lorraine.fr (J. Epstein).

https://doi.org/10.1016/j.jclinepi.2024.111488

What is new?

Key findings

- The study is dedicated to the development of an instrument targeting the flare phenomenon from the patient's perspective.

What this adds to what is known?

- The application of Rasch analysis in both subsamples (derivation and validation) led to a refined scalable version of a questionnaire to measure flare in knee or hip osteoarthritis.

What is the implication and what should change now?

- The strength of the 16-item Flare-OA questionnaire lies in its development, validity testing, and scalability assessment conducted in a dual-language approach with a multicultural sample.

of this phenomenon, such as flare, may aid in guiding treatment. In fact, a feasible diagnosis of flare is an important clinical issue. For this, similar to flare in other diseases such as chronic obstructive pulmonary disease [9,10], Gout [11] or asthma [12,13], an accurate definition for flare in OA was necessary [3,8,14]. The Outcome Measures in Rheumatology (OMERACT) group proposed that flare is defined as a transient state of the condition, with a duration of a few days, characterized by onset or worsening of pain, swelling, stiffness, and with associated impact on sleep, activity, functioning, and mood, that can resolve spontaneously or require adjusting therapy, even if only temporarily [15]. It was operationalized by considering 5 domains [3,8,14] which were endorsed by an international consensus of patients, scientists, and clinicians [16]: Pain, Swelling, Stiffness, Psychological aspects, and Impact of symptoms during flare. One of the landmarks of this group work was to consider that flare occurs mostly outside the scope of clinical presentation. Therefore, measurement from the patient perspective is crucial to account for the experience and consequences of a flare, to search for predictors of occurrence, and to assess the effectiveness of flare treatments.

A measurement instrument has been developed using a dual-language approach (French and English) following a mixed-method study in an international setting involving OA patients, clinicians, and scientists. Steps to its development and validation were taken under the umbrella of the OMERACT organization [17] that follows the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) taxonomy for measurement properties [18]. In the COSMIN initiative, the quality of properties instruments relevant for evaluating patient-reported outcome measures are evaluated in terms of reliability, validity, and responsiveness. The newly developed Flare-OA self-report questionnaire measuring flare in knee and hip OA in the patient perspective (19 items on a numerical rating scale in 5 domains) showed good psychometric properties assessed along the classical test theory [19]. The COSMIN quality checklist also pointed on the interest of the item response theory approach as a complementary method for assessing scale validity [18,20]. The aim of the present study was to determine the Flare-OA scaling properties by Rasch analysis and to present evidence for a refined scalable version for optimal use in clinical research practice.

2. Materials and methods

The study included patients recruited from the Royal North Shore Hospital, Sydney, Australia, the knee and hip osteoarthritis long-term assessment cohort followed in 6 rheumatology centers and 4 other rheumatology centers in France, and the OA Action Alliance in the United States, a network of self-reported OA patients. Inclusion criteria were at least 45 years of age, with symptomatic clinical and radiological knee or hip OA, confirmed by a physician (except for the United States where there was no such confirmation of diagnosis) and who had completed an online questionnaire in English or French. Exclusion criterion was having both hip and knee OA. The participants were 398 subjects with mean age 64 years (standard deviation = 8.1), 70.4% women, and 86.7% had knee OA, and 60 patients answered all questionnaire items and self-reported flare 14 days later (53 and 7 with knee and hip OA, respectively).

2.1. Instrument

The Flare-OA questionnaire is a self-reported assessment of flare occurrence over the past 4 weeks. The patients use an 11-point numerical rating scale with a response indicating to what extent he/she agrees with each statement (0 = Not at all to 10 = Absolutely). The Flare-OA questionnaire consists of 19 statements assessing 5 domains as delineated in flare definition, that is, Pain, Swelling, Stiffness, Psychological aspects, and Impact of symptoms.

2.2. Data analysis

To validate the scaling properties of the instrument, Rasch model analysis was conducted using RUMM2030 software (RUMM Laboratory Pty, Ltd, Duncraig, WA). The sample was split randomly in a *derivation* and a *validation* sample stratified by country and by type of affected joint, with a 2:1 split ratio. The Rasch analysis was an

iterative and dynamic process conducted in steps that could be didactically described in 8 steps [21].

In step 1, all items for each of 4 domains were fitted to a Rasch model tested in the derivation sample. The dimension Swelling with only one item was not eligible for Rasch modeling. Appropriate Rasch model was chosen using a likelihood ratio test. A summary test-of-fit statistic was derived.

In step 2, items with disordered thresholds, that is, with some person higher ability (trait) reflected in lower—instead of higher—categories of measurement response option, were identified from the threshold map and were reordered by collapsing adjacent categories using item category probability curve.

In step 3, the model was retested for goodness-of-fit. Fit was determined when an item-trait interaction chi-square *P* value was nonsignificant with a Bonferroni adjusted correction for number of items used and by item-fit and person-fit residuals. Reliability was evaluated with Person Separation Index (PSI), where a value of 0.70 or more is required for acceptable reliability and a value of 0.61 for minimal reliability (separation between 2 groups [22]). Individual item-fit residual should be within the ±2.5 range, with a nonsignificant associated chi-square *P* value. When needed, items with a misfit were deleted.

In step 4, the Rasch model for each dimension was reanalyzed to re-evaluate and validate goodness-of fit.

In step 5, each item was also examined for differential item functioning (DIF) considering sex, age (<60, 60−70, >70), country (Australia, the United States, France), and joint (knee, hip), graphically and statistically (analysis of variance). A significant uniform DIF for an item was solved by splitting an item by different levels of groups where DIF occurred. Items with a significant nonuniform DIF were deleted.

In step 6, unidimensionality assumption was checked with the test proposed by Smith [23]. Through principal components analysis of residuals, person-locations were then compared for each person using an independent *t*-test based on 2 subsets of items with the highest positive and the highest negative loadings on the first principal component. If less than 5% of *t*-test comparisons were significant, this was considered evidence for unidimensionality.

In step 7, local dependency between items was detected when a residual correlation was 0.2 point above the average of all items residual correlations (24). Local dependency was solved by combining items modalities into a superitem, a solution which involved a sum of the dependent items also called ‘‘subtest’’ [24].

In step 8, person-item threshold distribution which displays the distribution of the person locations (above x-axis) and distribution of item threshold locations (below x-axis) on the same scale along the trait was examined to access scale targeting. After building each model in the derivation sample, we applied it to the validation sample to confirm the instrument properties, checking each of 8 steps described above. In case of persistent reversed threshold in validation sample, rescoring of items appropriate to validation sample was realized and Rasch analysis for associated dimension recomputed on both samples.

Following recommendations of COSMIN initiative [18] and to reproduce what was performed on the full version [19], the psychometric properties of the resulting reduced questionnaire were documented as follows: (1) the structural validity, that is, the degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured, was assessed with a confirmatory factor analysis (CFA) with maximum likelihood method estimation, a method to assess the adequation of the items with the prespecified 5-dimension model, using goodness-of-fit indices: the comparative fit index (≥0.95), the Tucker Lewis index (≥0.95) [25], the standardized root mean square residual (≤0.08), and the root mean square error of approximation (≤0.06) and (2) the construct validity, that is, the degree to which the scores of an instrument are consistent with hypotheses (namely the relationships to scores of other instruments) based on the assumption that the instrument validly measures the construct to be measured, was assessed by the correlation with the knee injury and osteoarthritis outcome score (KOOS), hip injury and osteoarthritis outcome score (HOOS), and Mini−Osteoarthritis Knee and Hip Quality of Life (Mini-OAKHQOL). The intrascore test-retest reliability at 14 days was assessed by intraclass correlation coefficient and the minimal detectable change was estimated with the Standard Error Measurement (SEM) calculation. The CFA and the correlation analyses were conducted with the Statistical Analysis Software (SAS), version 9.4.

3. Results

Rasch analysis was conducted with 167 to 203 subjects in the derivation sample (of 266) and with 90 to 105 subjects in the validation sample (of 132), due to extreme scores or missing data per item.

3.1. Choice of the Rasch model

The likelihood ratio tests conducted with a derivation subsample on 4 domains were significant ($P < .001$), which rejected the rating scale model and indicated the appropriateness of the partial credit Rasch model for the current analysis.

The initial Rasch analysis indicated satisfactory reliability PSI for all domains (>0.70). But the overall initial model fit was poor for 3 domains ($P < .01$), and 16 items displayed disordered thresholds. The stiffness dimension was the only one that showed no need for recoding ($P = .45$). A recoding procedure was applied to all items of the other domains.

3.2. Item rescoring

The items showed disordered thresholds with different patterns. Optimal ordering of thresholds was achieved through nonuniform rescoring of all 11-point scale responses through collapsing response categories to a shorter-point scale. Figure 1 provides an example of a typical item category probability curve, displaying a disordered threshold (left) and threshold ordered (right) after final rescoring, respectively.

Item 11 (*I felt more restricted or impaired in my movements*) in dimension Consequence of symptoms presented misfit (fit residual = −2.583; P = .0055), and item 15 (*I felt frustrated because I was limited in my daily activities*) in dimension Psychological aspects (fit residual = −2.473; P = .0041) led to discarding these items.

3.3. Unidimensionality

We observed that less than 5% of the *t*-tests were significant, confirming the unidimensionality of scale in 4 domains.

3.4. Local dependency

The residual correlation matrix indicated local dependency between item 1 (*My pain felt more severe compared to my usual pain*) and item 3 (*My pain was more persistent than usual*), which was solved by creating a super-item score combining response modalities of both items into one scale. Good model fit was found on re-evaluation and no further local dependency was confirmed.

3.5. Differential item functioning

A uniform DIF was detected for joint in item 4 (*My pain disrupted my sleep more than usual*) in dimension Pain (Fig 2, on the top). As the observed DIF was clinically consistent with clinical difference between knee and hip symptoms, we decided to retain the item in its dimension and solved DIF problem by splitting item 4 across joint.

Item 9 (*I needed to put ice or something cold on my joint more than usual*) presented a nonuniform DIF according to the age (Fig 2, on the left) and the country (Fig 2, on the right) and was excluded from its dimension Consequence of symptoms.

3.6. Person-item threshold distribution

The person-item threshold distribution of the rescored 16-item Flare-OA (Flare-OA-16) questionnaire is shown in Figure 3 for both the derivation and validation samples.

Overall, there was a good match between location of items and of persons over each dimension trait in derivation and validation samples, with a regular spread of thresholds despite some gaps inside trait. This aspect is particularly important in clinical evaluation, that is, covering the different aspects of the latent trait predicted for the target population. In addition, the mean location for persons was close to 0 (from −0.478 to 0.416), and domains of Flare-OA-16 were well focused, that is, not too *easy* and not too *hard* for the targeted population.

3.7. Final model

A new round of analysis was performed in the derivation sample and in the validation sample. Modifications decided at previous steps allowed improvement of the overall fit model for 4 domains (P > .05, Table 1) in both samples, with acceptable reliability in 3 domains (PSI > 0.7) and minimal reliability in the Pain dimension (PSI = 0.62 and 0.64) (Table 1).

The threshold map for all domains of the resulting Flare-OA-16 questionnaire is shown in Figure 4 for derivation and validation samples.

In CFA results (maximum likelihood method), factors with at least 2 items were entered and showed good fit indicators: comparative fit index = 0.98; Tucker Lewis index = 0.97; standardized root mean square residual = 0.04; and root mean square error of approximation = 0.06.

Considering the construct validity analysis, results are disclosed in Table 2.

Coefficients were from 0.61 to 0.86 for the Flare-OA score[1] correlated with scores for pain, symptoms, and ADL dimensions of the HOOS and KOOS and pain, physical activities, and mental health dimensions of the Mini-OAKHQOL. For the other dimensions of the HOOS and KOOS, coefficients ranged from 0.52 to 0.69. For the other dimensions of the Mini-OAKHQOL, coefficients ranged from 0.12 to 0.63.

The reproducibility at 14 days was characterized with intraclass correlation coefficient = 0.84 (0.75−0.90) and the minimal detectable change was documented with a SEM = 1.09.

4. Discussion

The selection of the Flare-OA questionnaire as an instrument that effectively assesses flare-OA from the patient's perspective was carefully conducted on the basis of the OMERACT guide [17], which applies the COSMIN taxonomy [18,20]. In this article, we go further in refining the questionnaire and evaluating its psychometric properties. The application of Rasch analysis in both subsamples (derivation and validation) led to a refined scalable version of a questionnaire to measure flare in knee or hip OA of which 3 items were discarded: 2 due to misfit and 1 due to country DIF. The quality of the indicators

[1] In this case, the global score is a comprehensive representation of the measured construct. In the supplementary material, the correlations between the Flare-AO-16's domains and the analyzed instruments' dimensions can be observed. Those are very close to the ones presented for the overall score.

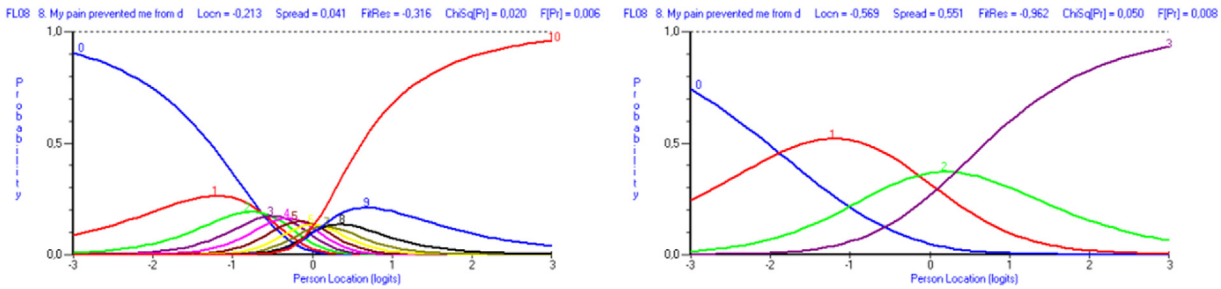

Figure 1. Item category probability curves for Item 8 of dimension Consequences of Symptoms, before rescoring (left) and after rescoring (right). Item fit statistics. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

obtained for Flare-OA-16 questionnaire was achieved through careful measures taken during application of the Rasch model.

In following recommendations from the literature [21,26], we improved the fit of the items at each step and in the final models. The recoding of the items in the analysis process allowed significant adjustments to be achieved. The final PSI proved good instrument reliability except for Pain (PSI = 0.63) which was lower than acceptable (0.70) but above minimal reliability 0.61 [23]. Results found in the validation sample effectively validated the decisions adopted with the derivation sample. When comparing our results to other studies, we found theoretical consistency in endorsed domains related to flare in OA [8,16], and methodological coherence in using a self-reported questionnaire, which proved useful for measuring the construct [4,27].

Absence of local dependency between items, an assumption for Rasch analysis [24], was not fully met, likely because the nature of the construct measured by the Flare-OA-16 questionnaire favors the inter-relationship between items. The residual correlation matrix indicated local dependency between items 1 (*My pain felt more severe compared to my usual pain*) and 3 (*My pain was more persistent than usual*), which suggests that the scores on these items influence each other. Item suppression attempts negatively affected the overall fit of the pain dimension, which confirm the importance of both items to measure the construct. Situations such as this are not uncommon in the literature of clinical instrument validation, which applied the Rasch model [15,22,27] because pairs of symptoms may be expected in clinical practice. For Flare-OA-16 questionnaire, a super-item was created for score calculation to solve this.

The clinical coherence was also considered for observed DIF for item 4 (*My pain disrupted my sleep more than usual*). The difference between knee and hip symptoms for this item was clinically consistent and the decision was to split item 4 across joint, similar to what was conducted in

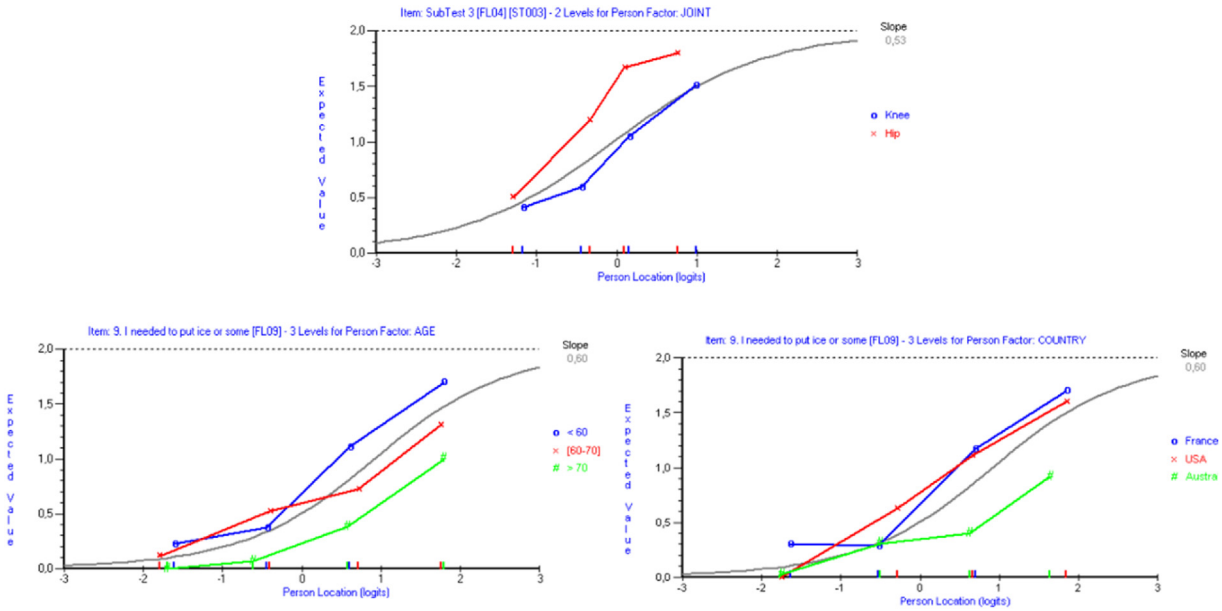

Figure 2. DIF for Item 4, dimension Pain, according to the joint (top) and DIF for Item 9, dimension Consequences of Symptoms, according to the age (left) and the country (right). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Figure 3. Person-item threshold distribution in derivation subsample (left) and validation subsample (right) by 4 domains. Note: Person's and item difficulty are plotted on the same interval logit scale. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

situations where DIF was presented in the item to measure intermittent pain [28].

In the fourth dimension, Psychological aspects, the analysis showed misfit for both item 15 (*I felt frustrated because I was limited in my daily activities*) and item 17 (*I needed to rest [eg, lie down or sit] to prevent my pain*). The evaluations previously performed by healthcare professionals and patients, published in a previous study [16], helped to achieve the final decision to remove item 15 that received low priority.

After decisions driven by Rasch analysis, the Flare-OA-16 version showed satisfactory construct validity (according to CFA results), as well as convergent validity (correlations observed with the HOOS, KOOS, and Mini-OAKHQOL) measures. However, some limitations should be mentioned. On one side, to obtain a single instrument measuring flare in knee and hip with accuracy, it was necessary to exclude patients affected with OA in both joints ($N = 13$). So, these patients are not represented in this study, but the content of the questionnaire was designed to be neutral to the affected joint (except for one item). Another limiting point concerns the swelling dimension that could not be included in Rasch analysis. To maintain consistency with the contributions of patients and experts to content validity, this dimension with only one item was kept in the final instrument.

Table 1. Summary test-of-fit statistics for last rescored items and after items deleted with derivation sample and validation sample

	Item fit residual		Person fit residual		Goodness of fit		Pearson separation item (PSI)	Independent *t*-test	
Dimension	**Value**	**SD**	**Value**	**SD**	**Chi-square (df)**	***P***		**%**	**95% CI**
Derivation sample									
Pain	0.267	1.613	−0.397	1.249	19.747	.072	0.62	NA[b]	0.267
Stiffness	−0.255	0.494	−0.592	0.897	8.245	.221	0.80	1.13	−0.255
Consequences of symptoms	0.013	0.826	−0.288	0.842	25.100	.014[a]	0.74	1.17	0.013
Psychological aspects	−0.176	1.389	−0.252	0.808	21.677	.117	0.75	1.57	−0.176
Validation sample									
Pain	0.234	1.180	−0.326	1.026	4.163	.842	0.64	NA[b]	0.234
Stiffness	0.028	0.467	−0.767	1.259	4.311	.366	0.73	1.53	0.028
Consequences of symptoms	0.093	0.613	−0.300	0.847	10.232	.249	0.77	1.56	0.093
Psychological aspects	0.030	1.173	−0.270	0.921	10.138	.428	0.78	0.00	0.030

[a] *p* adjustment de Bonferroni = 0.05/4 items final model = 0.0125.
[b] Not Applicable (split of item 4).

Among the strengths of this study, it is important to highlight that this is the first study of a dedicated measurement instrument targeting the flare phenomenon from the patient's perspective. The match between location of items and of persons over each dimension indicated the questionnaire covers the different aspects of flare-OA, which is particularly important in clinical evaluation. Ordinarily, to better detect the occurrence of OA flares and to adapt the short-term and long-term treatment, baseline indicators of pain intensity as well as intercorrelated aspects (such as anxiety, fatigue, and activity limitations) are taken into consideration [8,24,29]. But due to the difficulty in identifying a criterion indicating the duration and the end of a flare episode, patients' report could better describe the flare state. The other positive point of the Flare-OA-16 questionnaire lies in its development, validity testing, and scalability assessment conducted in a dual-language approach with a multicultural sample. The development process within a COSMIN taxonomy, endorsed by OMERACT, a broad scientific community of physicians, scientists, and patient partners, allowed to select relevant items. The psychometric properties assessed for the Flare-OA-16 were similar to those found for the full version, showing similar dimensionality and relevant correlation with instruments

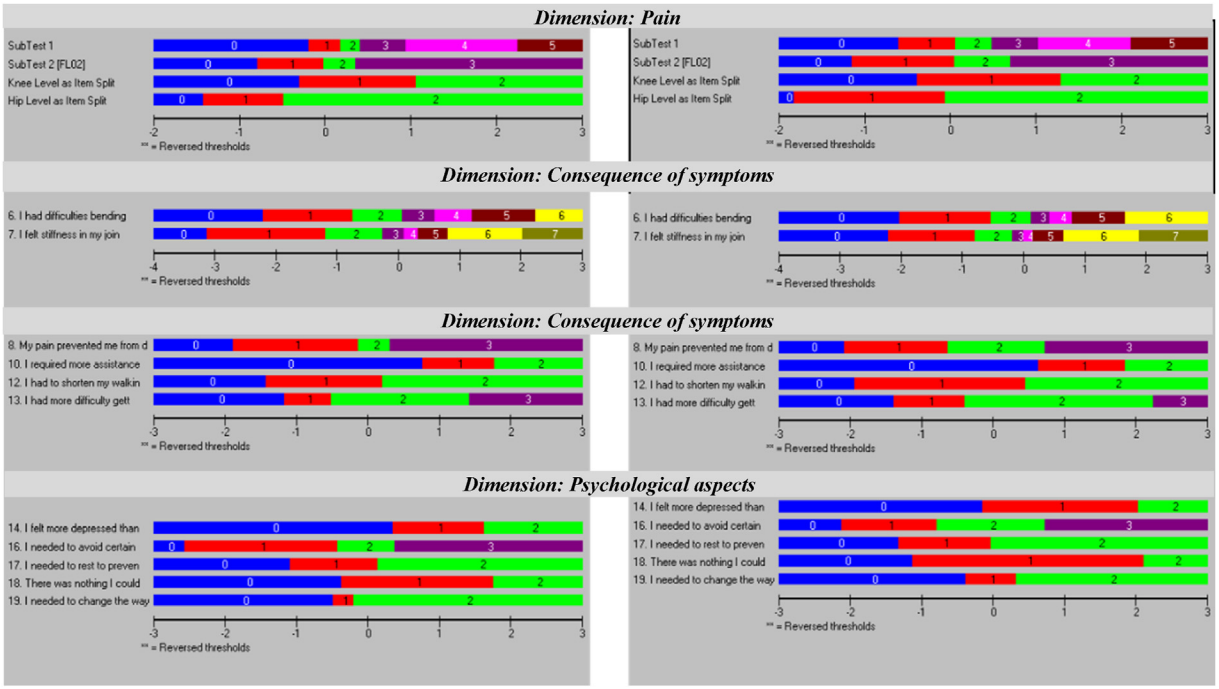

Figure 4. Threshold map for the Flare-OA items after nonuniform rescoring, ordered by difficulty, in derivation subsample (left) and validation subsample (right). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2. Pearson correlation between Flare-OA score and scores for functional limitations on the HOOS, KOOS, and Mini-OAKHQOL (*n* in parentheses indicates the number of respondents)

Validity	Mean (SD)	Correlation (r)
HOOS		
Pain (*n* = 46)	61.30 (23.53)	−0.86
Symptoms (*n* = 46)	61.52 (20.97)	−0.70
Function in daily living (ADL) (*n* = 44)	65.61 (23.04)	−0.81
Function in sport and recreation (Sport/Rec) (*n* = 44)	50.99 (28.88)	−0.69
Quality of life (QoL) (*n* = 44)	46.31 (26.04)	−0.71
KOOS		
Pain (*n* = 333)	57.57 (19.33)	−0.73
Symptoms (*n* = 333)	54.56 (20.05)	−0.61
Function in daily living (ADL) (*n* = 328)	62.86 (21.13)	−0.68
Function in sport and recreation (Sport/Rec) (*n* = 328)	31.53 (26.07)	−0.52
Quality of life (QoL) (*n* = 328)	39.83 (22.94)	−0.64
Mini-OAKHQOL		
Pain (*n* = 368)	53.66 (25.87)	−0.75
Physical activities (*n* = 368)	56.25 (27.58)	−0.73
Mental health (*n* = 368)	73.97 (27.19)	−0.69
Other dimensions		
Social support (*n* = 367)	56.10 (28.20)	−0.12
Social activities (*n* = 368)	67.73 (28.73)	−0.31
Professional activity (*n* = 365)	72.71 (31.54)	−0.55
Fear of dependent (*n* = 367)	66.05 (37.73)	−0.63
Sexual relation (*n* = 367)	76.59 (33.89)	−0.40

HOOS, hip injury and osteoarthritis outcome score; KOOS, knee injury and osteoarthritis outcome score; Mini-OAKHQOL, mini—Osteoarthritis Knee and Hip Quality of Life; SD, standard deviation.

measuring close constructs [19]. Applying the instrument in different countries with different cultures allowed us to warrant the communality of the construct and its suitability to these cultures. With good cross-cultural validity, extension and adaptation to other cultures should be easier to conduct. Studies are currently in progress to obtain responsiveness indicators (a crucial step suggested in the COSMIN taxonomy) beyond the current documentation of the SEM, that allows sample size calculation, as well as for a cultural adaptation of the Flare-OA questionnaire in other languages.

In practice, the Flare-OA-16 questionnaire is a self-reported questionnaire designed to explore the occurrence and severity of flare over the past 4 weeks, that can be used at several steps of an OA trial, for example, eligibility visit to identify active disease, inclusion visit to document flare-up, and follow-up visits (at a minimum of 4-week intervals) to measure flare outcome. The Rasch model analysis made it possible to have a shortened version that is useful for clinical research, yet to be explored for routine practice. Flare occurs mainly outside the clinical scope presentation. Therefore, having a reliable measure to analyze these episodes from the patient's perspective seems crucial. Taking into account the experience and consequences of flare is important to look for predictors and evaluate the effectiveness of treatments. To better achieve this, a global composite score should be better explored. Based on proposed taxonomy on this subject [30], some preliminary study indicates that it is possible to group 5-domain scores of Flare-OA-16 [31] into a composite score. Finally, the responsiveness property remains to be fully documented from ongoing studies.

5. Conclusion

In conclusion, we affirm the Flare-OA-16 questionnaire measures 5 domains conceptually consolidated for flare in OA [19]. The present study conducted in a multicultural sample has demonstrated that the instrument achieved expectations of the Rasch model after some modifications. The 16-item questionnaire has adequate psychometric properties and scalability for assessment of flare in hip and knee OA, according to the endorsed domains of the latent trait. In terms of feasibility, we confirmed that expert analysis (physicians, clinical researchers, and patients) had already indicated positive evaluations of the Flare-OA-16 questionnaire in terms of ease of use, cost, and effectiveness of assessment of the measured construct [32].

The Rasch analysis was helpful in guiding the decision to refine the measurement instrument making the Flare-OA-16 questionnaire available for use in clinical research.

Ethics statement

The National Commission for Data Protection in France (CNIL DR-2015-134) and the Ethics Committee on Human Research at the University of Sydney, Australia (2015/020) approved ethical and regulatory aspects. All patients gave informed consent to participate in the research (registered at ClinicalTrials.gov: NCT02892058).

CRediT authorship contribution statement

Fabiana Queiroga: Writing − review & editing, Writing − original draft, Validation, Methodology, Formal analysis, Conceptualization. **Jonathan Epstein:** Writing − review & editing, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Marie-Line Erpelding:** Writing − review & editing, Writing − original draft, Methodology, Formal analysis, Data curation. **Elisabeth Spitz:** Writing − review & editing, Validation. **Jean-Francis Maillefert:** Writing − review & editing, Methodology. **Bruno Fautrel:** Writing − review & editing, Data curation. **Leigh F. Callahan:** Writing − review & editing, Methodology. **David J. Hunter:** Writing − review & editing, Methodology. **Francis Guillemin:** Writing − review & editing, Writing − original draft, Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Data availability

Data are available on reasonable request to the scientific committee of the study.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors acknowledge the participants of the study samples who were used in the study, without whom this research could not have been possible. The authors are indebted to Dorcas Beaton for her analysis suggestions that improved the quality of the manuscript.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2024.111488.

References

[1] Cross M, Smith E, Hoy D, Guillemin F, Hill CL, Laslett LL, et al. The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014;73(7):1323−30.

[2] Safiri S, Kolahi AA, Smith E, Hill C, Bettampadi D, Mansournia MA, et al. Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. Ann Rheum Dis 2020;79(6):819−28.

[3] Cross M, Dubouis L, Mangin M, Hunter DJ, March L, Hawker G, et al. Defining flare in osteoarthritis of the hip and knee: a systematic literature review - OMERACT virtual special interest group. J Rheumatol 2017;44:1920−7.

[4] Thomas MJ, Butler-Walley S, Rathod-Mistry T, Mayson Z, Parry EL, Pope C, et al. Acute flares of knee osteoarthritis in primary care: a feasibility and pilot case-crossover study. Pilot Feasibility Stud 2018;4:167.

[5] Katz JN, Arant KR, Loeser RF. Diagnosis and treatment of hip and knee osteoarthritis: a review. JAMA 2021;325:568.

[6] Katz JN, Arant K, Loeser RF. Review of hip and knee osteoarthritis—reply. JAMA 2021;325:2505.

[7] Phillips RE. Review of hip and knee osteoarthritis. JAMA 2021;325:2504.

[8] Thomas MJ, Guillemin F, Neogi T. Osteoarthritis flares. Clin Geriatr Med 2022;38(2):239−57.

[9] Elvekjaer M, Carlsson C, Rasmussen S, Porsbjerg CM, Grønbæk KK, Haahr-Raunkjær C, et al. Agreement between wireless and standard measurements of vital signs in acute exacerbation of chronic obstructive pulmonary disease: a clinical validation study. Physiol Meas 2021;42(5).

[10] Singh D, Agusti A, Anzueto A, Barnes PJ, Bourbeau J, Celli BR, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease: the GOLD science committee report 2019. Eur Respir J 2019;53(5):1900164.

[11] Elmagboul N, Coburn B, Foster J, Mudano A, Melnick J, Bergman D, et al. Physical activity measured using wearable activity tracking devices associated with gout flares. Arthritis Res Ther 2020;22(1):181.

[12] Abe Y, Suzuki M, Kimura H, Shimizu K, Makita H, Nishimura M, et al. Annual fractional exhaled nitric oxide measurements and exacerbations in severe asthma. J Asthma Allergy 2020;13:731−41.

[13] Luckie K, Pang TC, Kritikos V, Saini B, Moles RJ. Systematic review and content analysis of asthma knowledge questionnaires: a focus on the knowledge surrounding acute exacerbation management. J Asthma 2018;55(6):615−28.

[14] Parry EL, Thomas MJ, Peat G. Defining acute flares in knee osteoarthritis: a systematic review. BMJ Open 2018;8(7):e019804.

[15] Guillemin F, Ricatte C, Barcenilla-Wong A, Schoumacker A, Cross M, Alleyrat C, et al. Developing a preliminary definition and domains of flare in knee and hip osteoarthritis (OA): consensus building of the flare-in-OA OMERACT group. J Rheumatol 2019;46:1188−91.

[16] King LK, Epstein J, Cross M, Buzzi M, Buttel T, Cembalo SM, et al. Endorsement of the domains of knee and hip osteoarthritis (OA) flare: a report from the OMERACT 2020 inaugural virtual consensus vote from the flares in OA working group. Semin Arthritis Rheum 2021;51(3):618−22.

[17] Beaton D, Maxwell LJ, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Chapter 5: instrument selection for core outcome measurement sets. *The OMERACT Handbook*. Filter 2.1 Updated April 1st 2021. The OMERACT Handbook © 2021 by OMERACT is licensed under

CC BY-ND 4.0. 2021. Available at: https://omeracthandbook.org. Accessed May 16, 2022.

[18] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010; 63:737−45.

[19] Traore Y, Epstein J, Spitz E, March L, Maillefert JF, Rutherford C, et al. Development and validation of the Flare-OA questionnaire for measuring flare in knee and hip osteoarthritis. Osteoarthritis Cartilage 2022;30:689−96.

[20] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19:539−49.

[21] Siegert RJ, Tennant A, Turner-Stokes L. Rasch analysis of the Beck Depression Inventory-II in a neurological rehabilitation sample. Disabil Rehabil 2010;32(1):8−17.

[22] Penta M, Arnoul C, Decruynaere C. Développer et interpréter une échelle de mesure: Applications du modéle de Rasch. Primento Publishing; 2013. https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5693223. Accessed February 28, 2022.

[23] Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3(2):205−31.

[24] Andrich D. A rating formulation for ordered response categories. Psychometrika 1978;43(4):561−73.

[25] Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model Multidiscip J 1999;6(1):1−55.

[26] Sprague E, Siegert RJ, Medvedev O, Roberts MH. Rasch analysis of the edmonton symptom assessment system. J Pain Symptom Manage 2018;55:1356−63.

[27] Thomas MJ, Yu D, Nicholls E, Bierma-Zeinstra S, Conaghan PG, Stoner KJ, et al. Short-term recovery trajectories of acute flares in knee pain: a UK-Netherlands multicenter prospective cohort analysis. Arthritis Care Res 2020;72(12):1687−92.

[28] Moreton BJ, Wheeler M, Walsh DA, Lincoln NB. Rasch analysis of the intermittent and constant osteoarthritis pain (ICOAP) scale. Osteoarthritis Cartilage 2012;20:1109−15.

[29] Alageel M, Al Turki A, Alhandi A, Alohali R, Alsalem R, Aleissa S. Cross-cultural adaptation and validation of the Arabic version of the intermittent and constant osteoarthritis pain questionnaire. Sports Med Int Open 2020;4(1):E8−12.

[30] Wells GA, Tugwell P, Tomasson G, Guillemin F, Maxwell LJ, Shea BJ, et al. Composite outcomes at OMERACT: multi-outcome domains and composite outcome domains. Semin Arthritis Rheum 2021;51(6):1370−7.

[31] Queiroga F, Epstein J, Erpelding ML, Soudant M, King L, Spitz E, et al. Building a composite score for patient self-report of flare in osteoarthritis: a comparison of methods with the Flare-OA-16 questionnaire. J Clin Epidemiol 2024:111467.

[32] Queiroga F, Cross M, Thomas MJ, March L, Epstein J, Guillemin F. A scoping review of patient self-report measures of flare in knee and hip osteoarthritis (OA): a report from the OMERACT flares in OA working group. Semin Arthritis Rheum 2023;63:152281.