Multi-group invariance testing of the knee injury osteoarthritis outcome score for joint replacement scale

Caleb Allred, Ashley J. Reeves, Madeline P. Casanova, Adam C. Cady, Russell T. Baker

ARTICLE INFO

Keywords:
Structural validity
Psychometric analysis
Knee pathology
Total knee arthroplasty

ABSTRACT

Objective: The Knee Osteoarthritis Outcome Score for Joint Replacement (KOOS-JR) scale is commonly used to assess patient progress. Scale structural validity has not been completely assessed. The purpose of this study was to assess the internal consistency, structural validity, and multi-group invariance properties of the KOOS-JR in a large sample of patients receiving knee arthroplasty or non-operative care.

Methods: A cross-sectional study using the Surgical Outcome System (SOS) database. Patients receiving care for degenerative knee conditions were included in the study. Internal consistency was assessed using Cronbach's alpha and McDonald's Omega. A confirmatory factor analysis was conducted to confirm scale structure of the KOOS-JR using a priori cut-off values (Comparative Fit Index [CFI], Tucker-Lewis Index [TLI], Incremental Fit Index [IFI]) ≥ 0.95, Root Mean Square Error of Approximation [RMSEA] ≤ 0.06 preferred and ≤ 0.08 acceptable). Multigroup invariance testing was conducted across sex, age, and intervention groups.

Results: Internal consistency was acceptable (alpha = 0.83; omega = 0.83). The unidimensional structure of the KOOS-JR exceeded most contemporary model fit recommendations (CFI = 0.976, TLI = 0.964, IFI = 0.976, RMSEA = 0.067). The KOOS-JR was invariant across groups, allowing for comparison of variances and means between sex, age, and intervention groups.

Conclusion: The KOOS-JR met or exceeded most of the recommendations for model fit. The scale can be used to assess differences between males and females, middle and older aged adults, and between baseline measures of patients who received total knee arthroplasty or non-operative care.

1. Introduction

Patient outcomes (e.g., pain, quality of life) are evaluated using various reporting method perspectives (e.g., clinician, self, physiologic) [1]. Although clinician and physiologic-reported outcomes are valuable, recent emphasis has been placed on understanding the patient's perspective of their injury or well-being using patient-reported outcome measures (PROMs) [2–4]. PROMs can positively inform patient care by providing valuable insight into clinical intervention effectiveness [5–8]. Thus, researchers have developed health related (e.g., quality of life, pain, disability) and joint-specific (e.g., knee, ankle) PROMs [5,7]. A commonly used unidimensional joint-specific PROM is the Knee Injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS-JR).

The KOOS-JR was developed using Rasch analysis on patient responses to the 42 items of the Knee Injury and Osteoarthritis Outcome Score (KOOS) instrument [9]. The KOOS was originally designed as a multi-dimensional scale for patients with end-stage osteoarthritis scheduled to undergo total knee arthroplasty [9]. The KOOS-JR uses seven of the original items to assess patient perceptions of overall knee health using questions pertaining to functional limitations, pain, and symptomology [9]; however, the KOOS-JR has been proposed as a unidimensional scale that provides a single score to quantify knee health [9]. The reduced length makes the KOOS-JR an attractive option for clinicians and patients alike (e.g., reduced response burden, reduced...
administration times, etc.) [10–14]. Researchers have found evidence of construct validity between the KOOS-JR and the KOOS Pain and KOOS Activities of Daily Living subscales [10]. Additionally, minimal detectable change (MDC) and minimal clinically important difference (MCID) values were calculated for the scale [11,14], and researchers have reported the scale demonstrated good responsiveness [10,12,13]. Thus, some initial findings provided evidence the scale can be used to measure the intended constructs and change in knee health across time [10,12,13]. However, subsequent psychometric evaluations of the scale have varied, particularly across statistical analysis approaches [11,14]. Researchers [15–17] have identified potential concerns with the psychometric properties (e.g., poor person differentiation, etc.) of the KOOS-JR, and examinations of the internal consistency, structural validity, and multi-group invariance testing of the scale are lacking. Specifically, there is a need to assess the internal consistency, structural validity and invariance of the KOOS-JR in a large, heterogeneous sample of participants who only responded to the KOOS-JR items [18–24]; prior assessment of the scale [9] used responses to all the KOOS items and it is possible that participant responses to the 7 retained items were influenced by the other KOOS items or the provided responses to those items not retained in the KOOS-JR.

Therefore, further analysis should be conducted to assess measurement properties of the KOOS-JR to guide scale use in clinical practice and research. Recommended scale development steps include assessing scale internal consistency (e.g., Cronbach’s alpha, etc.) and conducting confirmatory factor analysis (CFA) and multi-group invariance testing in a large, heterogeneous sample of patients who only respond to the 7-items in the KOOS-JR [18–24]. Evaluating internal consistency is important to support dimensional structure (i.e., support the claim of a unidimensional structure) and provide evidence that redundant or parallel items are not included in the scale [18–21]. Conducting CFA is necessary to assess the structural validity, which would fill a gap in the literature for confirming structural validity in a large heterogeneous sample who completed the KOOS-JR as a unique PRO. Further, multi-group invariance testing ensures the scale can be used to assess group differences and would indicate items are being interpreted similarly and the latent construct is being operationalized similarly across groups (e.g., sex, age, etc.) [22–24]. Thus, multi-group invariance would provide evidence that group differences are outside of measurement bias or error and would support scale use to assess group differences and test hypotheses [22,23,25]. Therefore, the purpose of this study was to assess the internal consistency, structural validity, and multi-group invariance properties of the KOOS-JR in a large sample of patients who would receive knee arthroplasty or non-operative care.

2. Methods

The Surgical Outcome System (SOS) is an international deidentified patient-reported outcome database that adheres to the Health Insurance Portability and Accountability Act (HIPAA), has already received Institutional Review Board (IRB) approval, and allows for retrospective analysis of previously collected patient data. The university IRB indicated approval for this study was not required because analysis of the deidentified data set from the SOS database was not considered human subject research. However, IRB approval was granted by the Cedar-Sinai Office of Research Compliance and Quality Improvement as part of a larger research project using SOS data. The data set utilized in this study included patients who had completed the KOOS-JR prior to receiving treatment/surgical intervention (i.e., knee arthroplasty, non-operative care) within the SOS database. Patient consent was obtained prior to the completion of patient-reported outcomes to be included in the SOS database.

2.1. KOOS-JR

The KOOS-JR is a 7-item instrument with questions pertaining to stiffness, pain, and knee function [9]. Patients respond to the seven questions using a 5-point Likert scale (none = 0, mild = 1, moderate = 2, severe = 3, extreme = 4) [26]. Raw scores are summed and range from 0 to 28, with 0 indicating perfect patient-perceived knee health. Raw scores are then transformed into an interval score, ranging from 0 (raw score of 28) to 100 (raw score of 0) [9]. For the purpose of this study, scores were converted to the transformed score.

2.2. Data analysis

KOOS-JR data and relevant demographic information were exported from the SOS database for analysis in the Statistical Package for Social Sciences (SPSS V. 25.0, Armonk, NY), the Hayes OMEGA SPSS package [27], and Analysis of Moment Structures (AMOS V. 25.0, Chicago, IL) software. Skewness values, kurtosis values, and histograms were assessed to evaluate normality of the data. Univariate and multivariate outliers were identified using z-scores (>3.3) and Mahalanobis distance, identified using a chi-square table with degrees of freedom and a p-value > 0.001 [23]. Cases in violation of the univariate or multivariate outlier criteria were removed. Descriptive statistics and frequencies were calculated for relevant demographic variables (e.g., age, sex) for the full sample and subgroups, and corresponding percentages were reported as appropriate.

2.3. Internal consistency

Internal consistency was assessed by calculating Cronbach’s alpha (α) and McDonald’s maximum likelihood omega (ω) for the 7-item unidimensional solution. Values < 0.70 indicate inadequate internal consistency, while values ≥ 0.90 may indicate item redundancy [20–23,28]. An acceptable range for Cronbach’s alpha (α) and McDonald’s maximum likelihood omega (ω) was set at ≥ 0.70 but ≤ 0.89 [20,28].

2.4. Scale structure – confirmatory factor analysis

A confirmatory factor analysis (CFA) with maximum likelihood estimation was conducted in AMOS on the proposed 7-item unidimensional KOOS-JR model. The model fit indices evaluated were set at established a priori values: Comparative Fit Index (CFI) ≥ 0.95; Tucker-Lewis Index (TLI) ≥ 0.95; Bollen Incremental Fit Index (IFI) ≥ 0.95; and root mean square error of approximation (RMSEA) ≤ 0.06 preferred and ≤ 0.08 acceptable [29,30]. The likelihood ratio statistic (CMIN) was calculated but was not used to assess model fit due to how heavily influenced the statistic is by sample size [22,23].

2.5. Multi-group invariance testing

A series of multi-group invariance analyses were conducted across relevant groups (i.e., intervention group [i.e., knee arthroplasty vs. non-operative care], sex [i.e., male, female], age group [youth: < 18, emerging adult: 18–25, early adulthood: 26–40, middle age: 41–65, older adult: > 65]) [31] to assess whether items were being interpreted equivalently across subgroups. Multi-group invariance testing was completed by testing three different models: configural model (i.e., to assess equal structure), metric model (i.e., to assess equal loadings), and scalar model (i.e., to assess equal intercepts), with each model being progressively more restrained than the previous model [25]. The CFI difference test (CFI_diff) and Chi-square difference test ($\chi^2_{\text{diff}}$ $\leq$ 0.01) were used to assess model fit. Model fit was considered adequate if CFI_diff was <0.01 [22,23,29]. While $\chi^2_{\text{diff}}$ was also assessed, the sensitivity of the statistic in large sample sizes led to $\chi^2_{\text{diff}}$ results being weighed less heavily in determining model fit [22,23,29]. Thus, multi-group invariance testing was continued if the CFI_diff criterion was met but the $\chi^2_{\text{diff}}$ criterion was violated.
3. Results

A total of 13470 complete responses (i.e., all items of the KOOS-JR were answered at baseline) from patients classified into a knee arthroplasty surgery group (n = 11564) or a non-operative care group (n = 1906) were extracted for data cleaning. From the sample of 13470 cases, 120 (knee arthroplasty: n = 110; non-operative: n = 10) univariate and multivariate cases were removed during the data cleaning process. A total of 1896 nonoperative cases and 11454 arthroplasty cases remained. To have an equal number of cases per group, a random sample of 1896 arthroplasty cases were extracted. The random selection of cases was generated in a three-step process: 1) a random number generator created the unique identifiers were then sorted in ascending order; 3) the first 1896 cases were selected. A final sample of 3792 total cases (female = 2127, 56.09%; male = 1527, 40.27%; not reported = 138, 3.64%; age = 63.31 ± 9.94 y [range = 12-89 y]) were used for analysis.

3.1. Internal consistency

Cronbach's alpha (α) and McDonald's maximum likelihood omega (ω) were conducted on the full sample of baseline scores for all items as a single dimension. Cronbach's alpha and McDonald's omega were calculated on the full sample of baseline scores for all items as a single dimension.

3.2. Scale structure - confirmatory factor analysis

The CFA conducted on the full sample (n = 3792) met most recommended model fit indices values (CFI = 0.976, TLI = 0.964, IFI = 0.976). The RMSEA value (0.067) exceeded the preferred fit criterion (≤0.06). The CFA conducted on the full sample met most recommended model fit indices values (CFI = 0.976, TLI = 0.964, IFI = 0.976). The RMSEA value (0.067) exceeded the preferred fit criterion (≤0.06).

![Fig. 1. Confirmatory Factor Analysis Measurement Model with Standardized Loadings. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; IFI = Bollen's Incremental Fit Index; RMSEA = Root Mean Square Error of Approximation, df = degrees of freedom. Each circle represents an unobserved variable, and each rectangle represents an observed variable. The unobserved variables to the right of each rectangle observed variable (e.g., e1, e2, e3, etc.) represent variance not explained by the factor that the indicator variable was intended to measure. The numbers along the lines from the KOOS-JR latent unobserved variable to each observed variable (e.g., Pre_1, Pre_2, etc.) are squared multiple correlations.](image)

3.3. Scale structure - exploratory factor analysis

The RMSEA value (0.067) exceeded the preferred cut-off value but met the criterion for acceptable RMSEA (Fig. 1). Modification indices did not demonstrate meaningful cross-loadings but indicated covariance between error terms for items 6 and 7; however, this was not specified as model fit was adequate.

4. Multi-group invariance testing

4.1. Demographic information is provided in Table 1

4.1.1. Intervention group

The full sample (n = 3792) was used for analysis. Most model fit indices were met for the baseline knee arthroplasty and non-operative care group models, except for RMSEA (knee arthroplasty = 0.072, non-operative care = 0.068). The initial multi-group model (i.e., configurural/equal form) exceeded all recommended fit indices (Table 2), and multi-group invariance testing continued. The metric model met fit criteria, warranting assessment of an equal latent variance model. The equal latent variance model passed all fit criteria, indicating variances were equal across groups. The scalar model also met model fit criteria, warranting examination of an equal latent means model. The equal latent means model did not pass the fit criteria, indicating the latent means were not equal across groups. Mean assessment indicated the non-operative care group reported significantly higher scores (i.e., better perception of knee health) than the knee arthroplasty group at baseline examination (Table 2).

4.1.2. Sex

A total of 3654 individuals (96.36%) reported their sex and were used for analysis (Table 1). The baseline models for sex met all preferred criteria except for RMSEA values; RMSEA values slightly exceeded the preferred cut-off value but met the criterion for adequate fit (Table 3). The initial multi-group model (i.e., configurual/equal form) exceeded all recommended fit indices; thus, multigroup invariance testing proceeded. The metric model met fit criteria, indicating assessment of an equal latent variance model was warranted. The equal latent variance model met fit criteria, indicating equal latent variances between groups. The scalar model also met fit criteria; thus, an equal latent means model was tested. The equal latent means model did not meet model fit criteria; further assessment of the means revealed females reported significantly lower scores (i.e., perceptions of worse knee health) than males (Table 3).

4.1.3. Age group

A total of 3612 (95.25%) cases from the middle age and older adult

| Table 1 Multigroup invariance testing group demographics. |
|------------------------------------------------------------|
| Characteristics                                      | N (%) | Mean Age (SD) | Males (%) | Females (%) |
|------------------------------------------------------------|
| Intervention Group                                       |       |               |           |             |
| Knee Arthroplasty                                         | 1896  | 65.56 ± 8.51   | 741 (40.90)| 1072 (59.10) |
| (50.00)                                                   |       |               |           |             |
| Non-Operative Care                                        |       |               |           |             |
| Characteristic                                           | N (%) | Mean Age (SD) | Arthroplasty (%) | Non-Op Care (%) |
|-----------------------------------------------------------|
| Sex Group                                                 |       |               |           |             |
| Males                                                     | 2127  | 63.51 ± 9.72   | 1072 (50.40)| 1055 (49.60) |
| (58.21)                                                   |       |               |           |             |
| Females                                                   | 1940  | 57.72 ± 5.71   | 837 (43.14)| 1103 (56.86) |
| (53.7)                                                    |       |               |           |             |
| Middle Age (41-65y)                                       | 1672  | 71.49 ± 4.58   | 996 (59.57)| 676 (40.43)  |
| (46.29)                                                   |       |               |           |             |
| Older Adult (+66y)                                        |       |               |           |             |

but met the criterion for acceptable RMSEA (Fig. 1). Modification indices did not demonstrate meaningful cross-loadings but indicated covariance between error terms for items 6 and 7; however, this was not specified as model fit was adequate.

| Age Group | Characteristics | N (%) | Mean Age (SD) | Arthroplasty (%) | Non-Op Care (%) |
|-----------|-----------------|-------|---------------|------------------|-----------------|
| Middle Age (41-65y) | 1672 | 71.49 ± 4.58 | 996 (59.57) | 676 (40.43) |
| Older Adult (+66y) | 1940 | 57.72 ± 5.71 | 837 (43.14) | 1103 (56.86) |
| Females | 2127 | 63.51 ± 9.72 | 1072 (50.40) | 1055 (49.60) |
| Males | 1896 | 65.56 ± 8.51 | 741 (40.90) | 1072 (59.10) |
| Intervention Group | 1896 | 65.56 ± 8.51 | 741 (40.90) | 1072 (59.10) |
The purpose of our study was to assess internal consistency and structural and multi-group invariance properties of the KOOS-JR in a large sample of patients with degenerative joint disease of the knee. Scale structure of the KOOS-JR was assessed using contemporary classical test theory procedures [23,29]. Our results support the structural validity of the KOOS-JR and provide evidence that the scale can be utilized to assess group differences based on sex, age, or intervention group (i.e., knee arthroplasty or non-operative care). Furthermore, acceptable internal consistency indicates scale parsimony and supports a unidimensional scale structure [20–23,28].

Our CFA results reveal sound model fit exceeding most of the recommendations for preferred model fit criteria. Thus, our findings further support prior Rasch analysis findings [9] of a structurally valid unidimensional model. We did not identify overall model fit concerns or local model fit concerns (e.g., low path coefficient loadings, meaningful item cross-loadings), that would suggest other alternative specifications to the scale are necessary to maximize fit or parsimony.

Our study also provides novel insight into the multi-group invariance properties of the scale to support and guide use of the KOOS-JR in clinical practice and research. Multi-group invariance testing supports scale hypothesis testing (e.g., can the scale be used to determine if patients with more severe injuries report higher levels of dysfunction), which provides valuable insight to clinicians [22,23,29]. Multigroup invariance testing also provides evidence that scale items are being interpreted similarly across groups (e.g., females, males) and that underlying constructs (i.e., knee health) are being measured similarly across groups. Thus,
Invariance allows for the comparison of scores across groups and indicates score differences in knee health are true group differences as opposed to differences in how group members interpret an item or operationalize the latent construct [22,23,29]. To our knowledge, we are the first group to present multi-group invariance testing results on the KOOS-JR.

We found the KOOS-JR was invariant at baseline measure (i.e., intake physical examination) between the two care intervention groups (i.e., knee arthroplasty and non-operative care). We found statically significant latent mean differences between the knee arthroplasty and non-operative care groups, with the non-surgical group reporting higher mean scores (i.e., higher level of perceived knee function) than the surgical group. While caution should be used in interpreting KOOS-JR scores as diagnostic, our findings provide preliminary support scale validity because patients with greater severity of knee degeneration (i.e., those who warrant surgical intervention) should report lower scores (i.e., greater knee health impairment) on the KOOS-JR. We would expect greater levels of perceived knee health impairment on the KOOS-JR to correlate with greater levels of joint degeneration (e.g., more advanced osteoarthritis) that would warrant surgical intervention. Our findings may provide support that baseline KOOS-JR scores function in this manner as the group means of patients who reported more impaired knee health on the KOOS-JR were the group who ultimately received surgical intervention.

However, other explanations (e.g., patients who seek conservative care may have better coping strategies for impaired knee health, etc.) could help explain the differences between treatment groups and warrant future research. Specifically, assessing baseline KOOS-JR scores across treatment groups in comparison to pathology (e.g., stage of OA, etc.) and patient psychosocial variables (e.g., coping strategies, resilience, quality of life, etc.) would be valuable for establishing scale validity and helping to determine if baseline KOOS-JR scores could support surgical or non-surgical intervention decisions for patients. Further analysis (e.g., longitudinal analysis) of the KOOS-JR would also be beneficial to determine if the KOOS-JR is invariant across time and if diagnostic cut-off criteria could be developed for the KOOS-JR to guide intervention decisions.

Our results also provide evidence that the KOOS-JR is invariant between groups of older adult populations (i.e., 41 years or older) and across sexes, which indicates the scale can be used to assess differences in knee health across these groups. Significant latent variance and latent mean differences were not found between the age groups suggesting minimal differences in knee health were perceived between the groups. The data set available, however, was not sufficient for multi-group testing of all age groups (i.e., 40 years or younger); thus, caution is warranted if assessment of group differences is conducted in these patient groups until further research is conducted to establish multi-group invariance and structural validity.

Statistically significant latent mean differences between the males and females in our sample were found. Females accounted for a larger proportion of the knee arthroplasty subgroup, which could partially explain these findings. For example, females were a larger proportion of the knee arthroplasty group, which may mean that more females presented with severe pathology, which would correlate with greater knee health impairment scores on the KOOS-JR compared to males. Other potential sex differences, however, may also help explain the latent mean sex differences given the percentage of females was similar in the knee arthroplasty (59.10%) and non-operative care (57.30%) groups. For example, men and women had their knee pain scores reported for condition prevalence, pain experiences, treatment responses, and age [32–35]. Researchers have also suggested women have reduced tolerance to painful stimuli [36] and a poorer capacity to cope with musculoskeletal pain due to higher levels of emotional distress and disability [37]. Others have indicated women report higher activity levels, pain acceptance, and social support than male counterparts who report more mood disturbances, higher kinesophobia, and lower activity levels with similar levels of perceived pain severity [35]. Thus, further research is needed to understand the underlying mechanisms for sex differences in knee health captured by the KOOS-JR.

While this study has several strengths, including a large, heterogeneous sample of patients seeking care, limitations do exist. First, all possible subgroups (e.g., younger populations, athletes, different surgical procedures) were not analyzed due to sample size limitations or the lack of information present in the SOS database. Thus, caution is warranted when examining KOOS-JR score differences in groups not analyzed. Further, the SOS dataset utilized did not include responses from healthy participants, nor did it include details on pathology (e.g., diagnosis, symptom duration), intervention utilized (e.g., specifics of non-operative care, surgical approach utilized), complete demographic information (e.g., ethnicity, physical activity level), or longitudinal data (e.g., baseline, post-surgery). Thus, many valuable analyses could not be performed, such as assessing test-retest reliability or calculating minimal detectable change and minimal clinically important differences (MCIDs).

Future research should determine if the KOOS-JR is invariant across younger age groups or across different levels of physical activity if the scale is to be used in those populations. Additionally, longitudinal invariance testing should be performed to ensure the measurement properties of the scale are maintained across repeated testing. Longitudinal testing also provides insight into if the KOOS-JR can be used to monitor recovery and guide patient care decisions (e.g., rehabilitation progression, discharge). Additionally, further longitudinal invariance and latent growth modeling analyses with more complete data (e.g., injury type, surgical approach, patient activity level) would allow for testing of substantive clinical questions (e.g., differences in treatment outcomes, rate of recovery between intervention techniques) that are important for guiding clinical practice.

6. Conclusion

The KOOS-JR met or exceeded most of the recommendations for model fit. Our findings support the structural validity of the KOOS-JR and the use of the scale to assess differences between males and females, middle and older aged adults, and between baseline measures of patients who received total knee arthroplasty or non-operative care. Further psychometric testing is necessary to establish reliability precision estimates (e.g., MCIDs) and longitudinal invariance to guide clinical use of the scale to assess patient progress over time.

Ethics approval and consent to participate

Institutional Review Board (IRB) approval for the project was granted by the Cedar-Sinai Office of Research Compliance and Quality Improvement as part of a larger research project using SOS data. University IRB was not required because the deidentified data set was not considered human subject research and the SOS adheres to the Health Insurance Portability and Accountability Act (HIPAA).

Availability of data and materials

The datasets analyzed during the study are not publicly available per study protocol; however, deidentified data may be available from the corresponding author with permission from Cedar-Sinai Office of Research Compliance and Quality Improvement, the Kerlan-Jobe Institute, and the University of Idaho upon reasonable request.

Competing interests

The authors declare they have no competing interests.

Funding

This publication was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of
the National Institutes of Health under Grant #P20GM103408 and an Idaho WWAMI Research Training Support Award.

Author contributions

CA – concept/design, data interpretation, manuscript drafting, revisions, final edits/approval. AJR – data analysis/interpretation, manuscript drafting, revisions, final edits/approval. MPC –data acquisition, data analysis/interpretation, manuscript drafting, revisions, final edits/approval. ACC – concept/design, data acquisition, manuscript drafting, revisions, final edits/approval. RTB – concept/design, data acquisition, data analysis/interpretation, manuscript drafting, revisions, final edits/approval.

Acknowledgements

Not applicable.

References

[1] R. Chin, B.Y. Lee, Principles and Practice of Clinical Trial Medicine, Elsevier Inc, London, UK, 2008.

[2] S. Cano, L. Pendrill, J. Melin, W. Fisher Jr., Towards consensus measurement standards for patient-centered outcomes, Measurement 141 (2019) 62–69.

[3] Institute of Medicine (US), Committee on quality of health care in America, in: Crossing The Quality Chasm A New Health System For the 21st Century, first ed., National Academies Press, Washington, DC, 2001.

[4] J.E. Ware Jr., M.K. Snyder, W.R. Wright, A.R. Davies, Defining and measuring patient satisfaction with medical care, Eval. Progr. Plann. 6 (3–4) (1983) 247–263.

[5] J.L. Farnsworth II, T. Evans, H. Binkley, M. Kang, Evaluation of knee-specific patient-reported outcome measures using Rasch analysis, J. Sport Rehabil. 30 (2) (2020) 278–285.

[6] A.M. Jette, Outcomes research: shifting the dominant research paradigm in physical therapy, Phys. Ther. 75 (11) (1995) 965–970.

[7] K.C. Lam, K.M. Harrington, K.L. Cameron, A.R. Snyder Valier, Use of patient-reported outcome measures (PROMs) for acute stroke: rationale, methods and considerations, and practical barriers, J. Athl. Train. 54 (4) (2019) 449–458.

[8] M. Reeves, L. Lisabeth, L. Williams, I. Katzan, M. Kapral, A. Deutsch, et al., Patient-reported outcome measures (PROMs) for acute stroke: rationale, methods and future directions, Stroke 49 (6) (2018) 1549–1556.

[9] S. Lyman, Y.Y. Lee, P.D. Franklin, W. Li, M.B. Cross, D.E. Fcheidt, Validation of the KOOS-JR: a short-form knee arthroplasty outcomes survey, Clin. Orthop. Relat. Res. 474 (6) (2016) 1461–1471.

[10] L.T. Hu, P.M. Bentler, Cutoff criteria for fit indexes in covariance structural analysis: conventional criteria versus new alternatives, Struct. Equ. Model. 6 (1) (1999) 1–55.

[11] C.K. Sigelman, E.A. Rider, Life-Span Human Development, ninth ed., Cengage Learning, Boston, MA, 2018.

[12] H. Breivik, B. Collett, V. Ventafridda, R. Cohen, D. Gallacher, Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment, Eur. J. Pain 10 (4) (2006) 287–333.

[13] B. Gerdle, J. Björklund, C. Henriksson, A. Bengtsson, Prevalence of current and chronic musculoskeletal pain, Int. J. Behav. Med. 7 (4) (2000) 305–321.

forms demonstrate responsiveness in patients undergoing knee arthroplasty, J. Arthroplasty 35 (6) (2020) 1540–1544.

[14] S. Lyman, Y.Y. Lee, A.S. McLawhorn, W. Islam, C.H. MacLean, What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? Clin. Orthop. Relat. Res. 476 (12) (2018) 2432–2441.

[15] J.L. Hunnicutt, B.N. Hand, C.M. Gregory, H.S. Stone, M.M. McLeod, B. Pietroboniome, et al., KOOS-JR demonstrates psychometric limitations in measuring knee health in individuals after ACL reconstruction, Sport Health 11 (3) (2019) 242–246.

[16] J.A. Padilla, H.L. Rudy, J.A. Gabor, S. Friedlander, R. Iorio, R.J. Karia, et al., Relationship between the patient-reported outcome measurement information system and traditional patient-reported outcomes for osteoarthritis, J. Arthroplasty 34 (2) (2019) 265–272.

[17] L. Eckhard, S. Munir, D. Wood, S. Talbot, R. Brighton, B. Walter, et al., The ceiling effects of patient reported outcome measures for total knee arthroplasty, Orthop. Traumatol. 107 (3) (2021), 102758.

[18] P. Panayides, Coefficient alpha: interpretation with caution, Eur. J. Psychol. 9 (4) (2013) 687–696.

[19] K. Pensdovs, J.M. Burr, C. Harley, D.B. Elliott, The development, assessment, and selection of questionnaires, Optom. Vis. Sci. 84 (8) (2007) 663–674.

[20] D.L. Streiner, Starting at the beginning: an introduction to coefficient alpha and internal consistency, J. Pers. Assess. 80 (1) (2003) 99–103.

[21] K.S. Taber, The use of Cronbach's alpha when developing and reporting research instruments in science education, Res. Sci. Educ. 48 (6) (2018) 1273–1296.

[22] T.A. Brown, Confirmatory Factor Analysis for Applied Research, second ed., Guilford Publications, New York, NY, 2014.

[23] R.B. Kline, Principles and Practice of Structural Equation Modeling, fourth ed., Guilford Press, New York, NY, 2016.

[24] L.B. Mokkink, C.A.C. Prinsen, D.L. Patrick, J. Alonso, H.C.W. de Vet, et al., COSMIN study design checklist for patient reported outcome measurement instruments, Available at: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-design-checklist_final.pdf. (Accessed 27 January 2022). Accessed.

[25] F.F. Chen, Sensitivity of goodness of fit indexes to lack of measurement invariance, Struct. Equ. Model. 14 (3) (2007) 464–504.

[26] E.M. Roos, L. Stefan Lohmander, The knee injury and osteoarthritis outcome score (KOOS): from joint injury to osteoarthritis, Health Qual. Life Outcome 1 (2003) 64.

[27] A.F. Hayes, J.J. Cous, Use omega rather than Cronbach's alpha for estimating reliability, But ... Commun. Methods Measur. 14 (1) (2020) 1–24.

[28] N.L. Leech, K.C. Barrett, G.A. Morgan, IBM SPSS for Intermediate Statistics: Use and Interpretation, fifth ed., Routledge, New York, NY, 2015.

[29] B.M. Byrne, Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming, third ed., Routledge, New York, NY, 2016.

[30] L.T. Hu, P.M. Bentler, Cutoff criteria for fit indexes in covariance structural analysis: conventional criteria versus new alternatives, Struct. Equ. Model. 6 (1) (1999) 1–55.

[31] C.K. Sigelman, E.A. Rider, Life-Span Human Development, ninth ed., Cengage Learning, Boston, MA, 2018.

[32] H. Breivik, B. Collett, V. Ventafridda, R. Cohen, D. Gallacher, Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment, Eur. J. Pain 10 (4) (2006) 287–333.

[33] B. Gerdle, J. Björklund, C. Henriksson, A. Bengtsson, Prevalence of current and chronic pain and their influences upon work and healthcare-seeking: a population study, J. Rheumatol. 31 (7) (2004) 1399–1406.

[34] D. Stubbbs, E. Kehs, M. Bair, T. Damba, J. Wu, J. Sutherland, et al., Sex differences in pain and pain-related disability among primary care patients with chronic musculoskeletal pain, Pain Med. 11 (2) (2010) 252–259.

[35] G.S. Rovner, K.S. Sunnerhagen, A. Bjorkdahl, B. Gerdle, B. Borsbo, F. Johansson, et al., Chronic pain and sex differences; women accept and move, while men feel blue, Pain Med. 11 (2) (2010) 232–244.

[36] G.S. Rovner, K.S. Sunnerhagen, A. Bjorkdahl, B. Gerdle, B. Borsbo, F. Johansson, et al., Sex differences and mechanisms of muscle pain, Curr. Opin. Pain. 11 (2019) 1–6.

[37] G. Grossi, J.F. Soares, U. Lundberg, Gender differences in coping with musculoskeletal pain, Int. J. Behav. Med. 7 (4) (2000) 305–321.