






Original Research

# Confirmatory Factor Analysis of the Hip Disability and Osteoarthritis Outcome Score (HOOS) and Associated Sub-scales

Emilie N. Miley<sup>1,2</sup> <sup>a</sup>, Madeline P. Casanova<sup>3</sup> , Scott W. Cheatham<sup>4</sup> , Lindsay Larkins<sup>4</sup> , Michael A. Pickering<sup>5</sup>, Russell T. Baker<sup>3</sup> 

<sup>1</sup> Department of Orthopaedic Surgery and Sports Medicine, University of Florida, <sup>2</sup> Department of Movement Sciences, University of Idaho, <sup>3</sup> WWAMI Medical Education Program, University of Idaho, <sup>4</sup> Movement Sciences, University of Idaho, <sup>5</sup> University of Idaho

Keywords: hip osteoarthritis, hip disability, physically active, patient outcomes

<https://doi.org/10.26603/001c.67938>

---

## International Journal of Sports Physical Therapy

Vol. 18, Issue 1, 2023

---

### Background

Hip Disability and Osteoarthritis Outcome Score (HOOS), HOOS-Joint Replacement (JR), HOOS Physical Function (PS), and HOOS-12 item scale have been suggested as reliable and valid instruments for assessing hip disability. However, factorial validity, invariance across subgroups, and repeated measures of the scale across different populations have not been well supported in the literature.

### Purpose

The primary study objectives were to: (1) assess model fit and psychometric properties of the original 40-item HOOS scale, (2) assess model fit of the HOOS-JR, (3) assess model fit of the HOOS-PS, and (4) assess model fit of the HOOS-12. A secondary objective was to perform multigroup invariance testing across physical activity level and hip pathology subgroups for models that met recommended fit indices.

### Study Design

Cross-Sectional Study

### Methods

Individual confirmatory factor analyses (CFAs) were conducted for the HOOS, HOOS-JR, HOOS-PS, and HOOS-12. Additionally, multigroup invariance testing (i.e., activity level, injury type) was conducted on the HOOS-JR and HOOS-PS.

### Results

Model fit indices did not meet contemporary recommendations for the HOOS and the HOOS-12. Model fit indices for the HOOS-JR and the HOOS-PS met some, but not all, contemporary recommendations. Invariance criteria was met for the HOOS-JR and HOOS-PS.

### Conclusion

The scale structure of the HOOS and HOOS-12 were not supported; however, preliminary evidence to support the scale structure of the HOOS-JR and HOOS-PS was found.

Clinicians and researchers who utilize the scales should do so with caution due to their limitations and untested properties until further research establishes the full psychometric properties of these scales and recommendations for their continued use.

---

**a Corresponding Author:**

Emilie N. Miley  
3450 Hull Road, Gainesville, FL 32607  
(352) 273-7361  
[mileyen@ortho.ufl.edu](mailto:mileyen@ortho.ufl.edu)

## INTRODUCTION

Hip osteoarthritis (OA) is a debilitating degenerative joint disorder that leads individuals to experience a multitude of symptoms including pain, disability in daily activity, reduced independence and quality of life (QoL).<sup>1,2</sup> With the multifaceted nature of hip OA on the rise, there is a need for a valid multidimensional (i.e., not specific to body location or injury) scale to adequately assess constructs across varied sub-populations.<sup>3-6</sup> Although several region-specific instruments (e.g., Lower Extremity Functional Scale) exist,<sup>3,4</sup> the majority of patient-reported outcome (PRO) measures for the hip joint primarily measure recovery following a total hip arthroplasty (THA) due to hip OA. Thus, many scales may not sufficiently assess all the relevant dimensions associated with OA (e.g., QoL), other pathologies, nor may be applicable to certain sub-populations (e.g., individuals who have not had a total hip replacement, younger active individuals, etc.). Additionally, some scales place an excessive response burden on patients and clinicians due to item redundancy, excessive number of items, or inclusion of items with inappropriate difficulty.<sup>7</sup> To adequately address these concerns, the Hip Disability and Osteoarthritis Score (HOOS) scale was developed.<sup>8-11</sup>

The HOOS consists of 40 items used to assess five dimensions: pain (10 items); other symptoms (five items); function in daily living (activities of daily living [ADL]; 17 items), function in sport and recreation (Sport/Rec; four items); and hip-related QoL (four items).<sup>7,10,12</sup> The HOOS can be used over both short-term and long-term intervals. For example, the HOOS can be used to evaluate changes from week-to-week, as produced by treatments such as medication, operation, or physical therapy, or to evaluate changes over years as a result of the primary injury or post-traumatic OA.<sup>7,12</sup> The HOOS is primarily intended to evaluate functional limitations and symptoms related to hip pathology or disability, with or without OA.<sup>7</sup> The HOOS has been studied in relatively small sample sizes ( $n < 200$ ) of adults aged 42-85,<sup>1,7,10,13</sup> and in patients who have either been diagnosed with hip OA or who have received a THA due to OA.<sup>7,14</sup> The HOOS, has not been extensively studied in healthy or younger populations, or in patients without hip OA. Additionally, the psychometric properties of the HOOS have not been established between groups (e.g., sex, different pathologies) or across patient visits (e.g., intake, discharge) using invariance testing.

In addition to the original 40-item HOOS, there have been several short-form versions created: the HOOS-JR (Joint Replacement), multiple HOOS-PS versions (Physical Function Short-Form), and the HOOS-12 item scale. The short-form versions were developed using items from different subdimensions of the original 40-item HOOS instrument, and have been studied in patients who have undergone THA.<sup>10,15,16</sup> The HOOS-JR includes six items from the original HOOS: two items from pain subscale, and four items from the function in daily living subscale.<sup>16</sup> One version of the HOOS-PS includes three items from the subscale function in daily living and two items from the Sport/Rec subscale<sup>10</sup>; whereas, other versions of the HOOS-PS in-

clude additional items (i.e., as many as seven items).<sup>15</sup> Because the HOOS-JR and HOOS-PS only provide a summary score and assess a single dimension,<sup>10,17</sup> the HOOS-12 short-form was developed to assess multiple dimensions.<sup>15,17</sup>

The HOOS-12 was created by developing an item bank using item response theory modeling, confirmatory factor analysis (CFA) procedures, and computerized adaptive test (CAT) simulations.<sup>17</sup> The item bank consisted of a set of items, taken from the original HOOS questions, that were identified to measure the same domain and parameters.<sup>17</sup> Individual CFAs were then performed on the individual constructs (i.e., pain, function, QoL) to verify that each item was unidimensional.<sup>17</sup> Following development of the item bank, CAT simulations were used to reduce the bank to include the fewest, yet most informative, items measuring each domain.<sup>17</sup> The final version of the HOOS-12 consisted of three constructs (pain, function, and QoL), that include four items from each of the subscales of the original HOOS.<sup>15,17</sup>

Psychometric examination of the HOOS and short-form versions have primarily focused on the construct validity, reliability, and responsiveness of the instrument. Construct validity has been established by correlating scores (i.e., Spearman's Correlation Coefficient) on the HOOS with the Short Form (SF)-36, which was intended to measure similar constructs within the HOOS (i.e., physical function vs. ADLs, physical function vs. sport and recreation, and bodily pain vs. pain), where low to moderate correlations ( $r = 0.49-0.66$ ) were found.<sup>7,18</sup> Lower correlations were identified among the HOOS and SF-36 constructs measuring mental health, whereas higher correlations were found between physical health constructs.<sup>7,19</sup> Internal consistency, or the assessment of homogeneity of the items, was assessed by interpreting Cronbach's alpha values, values ranging from  $> .70$  to  $\leq .90$  have been recommended.<sup>20-22</sup> For the HOOS, Cronbach's alpha values have ranged from 0.75 to 0.98 across multiple studies<sup>13,18,23</sup>; high values (i.e.,  $>.90$ ) may be indicative of potential issues (e.g., of item redundancy, construct underrepresentation, inclusion of too many items, etc.).<sup>21,24-26</sup> Test-retest reliability has also been assessed and values found have ranged from good to excellent (ICC = 0.75 to 0.97).<sup>13,18,23</sup> Finally, responsiveness to the 40-item HOOS has been assessed using the standardized response mean (SRM); researchers report a high response rate (SRM =  $> .80$ ) when compared to the Western Ontario and MacMaster Universities Osteoarthritis Index LK 3.0.<sup>7</sup>

The HOOS-JR and HOOS-12 have been reported to have acceptable internal consistency (0.70-0.92),<sup>15,16</sup> and high responsiveness (0.80).<sup>16</sup> External validity assessed using Spearman's Correlation Coefficient of the HOOS-JR has been reported to be acceptable with moderate to high correlations with the HOOS subscales (0.60-0.94)<sup>15,16</sup> and the HOOS-PS (0.81-0.86).<sup>16</sup> The HOOS-12 was also highly correlated with the HOOS ( $r = 0.75-0.94$ ).<sup>15</sup> The HOOS-JR, HOOS-PS, and HOOS-12 are all considered to be reliable (Cronbach's alpha = 0.77-0.92).<sup>15</sup>

Although the HOOS, HOOS-JR, HOOS-PS, and HOOS-12 have been suggested as reliable and valid instruments, factorial validity, invariance across subgroups, and repeated testing of the scales across different time points (i.e., longitudinal invariance) have not been well supported in the literature. Additionally, complete psychometric analysis of the HOOS and the short-form versions of the scale to ensure the instrument can be used in clinical practice and research have not been completed. Conducting a CFA to examine the factor structure of the proposed scales (i.e., HOOS, HOOS-JR, HOOS-PS, HOOS-12) and conducting CFA-based invariance testing to explore measurement properties of the scale across subgroups of the population (e.g., sex, physical activity levels, etc.), stages or types of musculoskeletal injury (i.e., healthy, acute, sub-acute, persistent, and chronic), and across time (i.e., intake, discharge) are warranted prior to adoption of a model for practice and research.<sup>26,27</sup> Establishing measurement properties through invariance testing ensures that the interpretations between groups or across time are valid and reliable.<sup>20,27</sup> Additionally, identifying a meaningful factor structure via CFA procedures enhances the rigor of psychometric examination of an instrument's measurement properties.<sup>20,27</sup>

Researchers have performed CFAs on the individual constructs (i.e., pain, function) proposed in the original HOOS<sup>17</sup>: some model fit recommendations for the pain and function constructs were met (CFI = 0.97-0.99, TLI = 0.97-0.98), while other construct fit indices did not meet recommendations (i.e., RMSEA = 0.14-0.19).<sup>17</sup> However, no assessment of the complete model structure, nor results from invariance testing have appeared in the literature to date. Thus, there exists need for additional assessment of the measurement properties of the HOOS and the proposed short form versions of the scale. Therefore, the primary purposes of this study were to: (1) assess the model fit of the original HOOS scale using a diverse sample to examine its psychometric properties, (2) assess the model fit of the HOOS-JR, (3) assess the model fit of the HOOS-PS, and (4) assess the model fit of the HOOS-12. The secondary purpose, if model fit held, was to perform multigroup invariance testing of the scale across physical activity level and hip pathology subgroups.

## METHODS

### PARTICIPANTS

After institutional review board approval (19-142), informed consent was obtained from all participants prior to data collection. Participants between the ages of 18 and 65 were recruited through social media (i.e., Facebook), email, and ResearchMatch. Participants self-reported their physical activity (i.e., inactive, low-, moderate-, high-activity), injury status (e.g., healthy, acute injury), and athlete level (e.g., competitive athlete, recreational athlete) classification (Table 1).

Social media recruitment was performed by providing study information (e.g., purpose of the study, inclusion/exclusion criteria, estimated survey duration) and a link to

the survey on multiple Facebook pages.<sup>28</sup> Email recruitment was performed by emailing the same study details from the social media recruitment and a separate survey link to a convenience sample of coaches (NCAA and recreational sports teams), athletic trainers, and ROTC officers across different higher education institutions, requesting the survey email to be shared with their athletes, patients, or cadets.

ResearchMatch was utilized as an additional online email recruitment tool. Searches of the ResearchMatch database were performed to identify potential participants (i.e., participants that were healthy, diagnosed with hip osteoarthritis, individuals who had undergone a THA, and those who had sustained a lower extremity musculoskeletal injury). Members of the database who fit the search criteria were recruited for this study per ResearchMatch protocol. Potential participants were identified and emailed invitations for study participation. If the participant voluntarily consented to participate, a follow-up email was sent to the individual containing a link to the survey.

### INSTRUMENTATION

Qualtrics (Qualtrics, LLC, Provo, UT) software was utilized to create an electronic survey via a weblink. The survey responses were collected directly into Qualtrics. Information collected included demographic information (e.g., age, sex, physical activity level) and responses to the items of the HOOS.

### HIP DISABILITY AND OSTEOARTHRITIS OUTCOME SCALE

The HOOS asks participants to rate how frequently they engaged in the behaviors over the past week using a 5-point Likert scale (1 = none/not at all/never, 2 = rarely/mild/monthly, 3 = sometimes/moderately/weekly, 4 = often/severe/daily, and 5 = extreme/always). Items were summed to create a score for each subscale, and global HOOS score, where 0 indicated extreme problems and 100 indicated no problems.<sup>7,12</sup>

### DATA ANALYSIS

Data were exported from the Qualtrics software and downloaded using Statistical Package for Social Sciences Version 24.0 (IBM Corp., Armonk, NY). Missing data were treated conservatively and any participant's data missing more than 10% of the responses on the HOOS (i.e., four or more missing responses) was removed from the data set.<sup>26</sup> Individuals missing less than 10% (i.e., three items or less) of the items within the HOOS were replaced with the mean score of the respective item for analysis purposes.<sup>26</sup> Participants with missing demographic data were not excluded from analysis and were left as missing values. Data were assessed for normality using z-scores, skewness, and kurtosis values. Multivariate outliers were also identified using descriptive statistics and Mahalanobis distance, the cut-off value was for 5 degrees of freedom at a p-value of 0.001, was 15.089.<sup>26,29</sup> This methodology generated the final data set used for analysis.

**Table 1. Terminology and Definitions**

Terminology	Definition
Physically Active <sup>4</sup>	"An individual who engages in athletic, recreational, or occupational activities that require physical skills and who uses strength, power, endurance, speed, flexibility, range of motion, or agility at least 3 days/week."
<b>Injury Classification<sup>4</sup></b>	
Healthy	"Free from musculoskeletal injury and fully able to participate in sport or activity."
Acute Injury	"A musculoskeletal injury that precludes full participation in sport or activity for at least 2 consecutive days (0–72 hours post-injury)."
Subacute Injury	"A musculoskeletal injury that precludes full participation in sport or activity for at least 2 consecutive days (3 days to 1-month post-injury)."
Persistent Pain	"A musculoskeletal injury that has been symptomatic for at least 1 month."
Chronic Pain	"Pain that consistently does not get any better with routine treatment or nonnarcotic medication."
<b>Athlete Level<sup>4</sup></b>	
Competitive athlete	"A participant who engages in a sport activity that requires at least 1 preparticipation examination, regular attendance at scheduled practices and/or conditioning sessions, and a coach who leads practices and/or competitions."
Recreational athlete	"A participant who meets the criteria for physical activity and participates in sport but does not meet the criteria for competitive status."
Occupational athlete	"A participant who meets the criteria for physical activity for occupation or recreation but does not meet the criteria."
Physically active in ADLs	"A participant who does not meet the criteria for any athlete category but who is physically active through daily activities (e.g., physically active for at least 30 min/day for 3 days/week)."

ADL = Activities of daily living

*SCALE STRUCTURE*

The final data set was used to conduct a CFA using Analysis of Moment Structures (AMOS) software (IBM Corp., Armonk, NY) on the 40-item HOOS and associated subscales. Consistent with the original proposed model, the HOOS scale was specified as a five factor, 40-item model.<sup>12</sup> Additionally, the HOOS-JR was specified as a one factor, 6-item model, the HOOS-PS was specified as a one factor, 5-item model, and the HOOS-12 was specified as a three factor, 12-item model. Given the subscales were created using items from the 40-item HOOS, the original item number labels were retained from the HOOS during the CFA procedures.<sup>10,15–17</sup> Full Information Maximum Likelihood Estimation was used to generate the parameter estimates. Model fit statistics included the likelihood ratio statistic (CMIN), Goodness of Fit Index (GFI), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Bollen’s Incremental Fit Index (IFI), and Root Mean Square Error of Approximation (RMSEA).<sup>26</sup> Model fit was evaluated based on *a priori* values: GFI ≥ 0.95, CFI ≥ 0.95, TLI ≥ 0.95, RMSEA ≤ 0.06, IFI; ≥ 0.95.<sup>26,30</sup> Latent construct correlations and path coefficient values with  $R^2 \geq 0.90$  were used to identify potential multicollinearity among the latent constructs, which indicates that item removal within a potential dimension might be beneficial to prevent model misspecification.<sup>31,32</sup>

*MULTIGROUP INVARIANCE TESTING*

Confirmatory factor analysis invariance testing was conducted if recommended model fit criteria were met to determine if the association between the latent constructs

(i.e., symptoms, pain, function, QoL) and the respective items were stable and equal across groups.<sup>26,31,33</sup> This was accomplished using a set of hierarchical procedures with an increasing level of constraint.<sup>26,31,33</sup> Individual CFAs were first conducted by subgroup category (i.e., activity level, injury type), ensuring the construct and factors (e.g., pain, function, symptoms) were measuring what was intended.<sup>31,33</sup> The model then underwent configural, metric, and scalar invariance testing.<sup>31–33</sup> First, the configural invariance test placed all groups in the same model to ensure the same factors have similar items across subgroups. Secondly, the metric model then tested if factor loadings were equal across subgroups.<sup>32</sup> If the model met metric invariance requirements, equal variances (i.e., group differences) between groups were then assessed.<sup>32</sup> Lastly, the scalar invariance test ensured that item intercepts were equal across groups, which indicated the means were not determined or altered by external factors.<sup>32</sup> If the model met scalar invariance requirements, equal mean models (i.e., score differences) were tested between groups.<sup>32</sup>

Model fit was compared using the CFI difference test (CFI<sub>DIFF</sub>) and the chi-square difference test ( $\chi^2_{DIFF}$ ), with a p-value cut-off of 0.01.<sup>30,33</sup> Given the sensitivity of the  $\chi^2_{DIFF}$  test to sample size,<sup>30</sup> the CFI<sub>DIFF</sub> test held greater weight in decisions regarding invariance testing model fit. If a model exceeded the  $\chi^2_{DIFF}$  test, but met the CFI<sub>DIFF</sub> test, invariance testing continued. Specifically, the HOOS-PS and HOOS-JR underwent invariance testing across physical activity level (i.e., inactive, low, medium, high) and injury type (i.e., hip OA and THA pathology, no hip pathology).

**Table 2. Demographics**

	Frequency (%) <sup>a</sup>
<b>Sex</b>	
Males	169 (25.8)
Females	481 (73.4)
<b>Injury Classification</b>	127 (19.4)
Healthy	453 (69.0)
Acute Injury	13 (2.0)
Subacute Injury	12 (2.0)
Persistent Injury	79 (15.0)
Chronic Injury	98 (15.0)
<b>Activity Level</b>	
Inactive	4 (6.1)
Low	225 (34.4)
Medium	276 (42.1)
High	113 (17.3)
<b>Athlete Level</b>	
Competitive athlete	31 (4.7)
Recreational athlete	198 (32.2)
Occupational athlete	127 (19.4)
Physically active in ADLs	122 (18.6)

<sup>a</sup> The sum does not equal 100% because percentages were rounded

**RESULTS**

Among the total responses (ResearchMatch = 487; social media = 370; total = 857), 149 participants were missing responses to more than 10% of the HOOS items and most of the demographic items. Thus, information could not be verified (e.g., sex, injury status) and these responses were removed from the dataset. Three individuals were missing responses to less than 10% of the HOOS; the missing values for those participants were replaced with the rounded mean for each item missing. Additionally, 53 (6.18%) participants reported scores that were identified as univariate ( $z$  scores  $\geq 3.4$ ) or multivariate (Mahalanobis distance  $\geq 15.089$ ) outliers and were removed from the dataset.<sup>29,34</sup> Of the participants removed, all injury categories (i.e., healthy, acute, persistent, chronic) and both sexes were represented. A total of 655 participants (i.e., social media/email:  $n = 247$  [37.7%]; ResearchMatch:  $n = 408$  [62.3%]) were included in the final data set (mean age =  $38.93 \pm 15.05$  yrs.; mean weight =  $165.42 \pm 41.99$  lbs.; [Table 2](#)). Participants self-reported their injury status and level of activity. The sample primarily included healthy participants (i.e., free of musculoskeletal injury;  $n = 453$ , 69%; [Table 2](#)), and the largest physically active response group indicated a level of moderate activity ( $n = 276$ , 42.1%; [Table 2](#)). Respondents reported participation in a variety of sports ([Table 3](#)) and a variety of injury locations ([Table 4](#)).

**Table 3. Participant – Reported Sport Activities**

Sport	Frequency (%) <sup>a</sup>
Track and Field	6 (0.9)
Basketball	2 (0.3)
Baseball	1 (0.2)
Volleyball	2 (0.3)
Soccer	8 (1.2)
Tennis	3 (0.5)
Golf	5 (0.5)
Swim and Dive	1 (0.2)
Rowing	60 (9.2)
Climbing	12 (1.8)
Rodeo	1 (0.2)
Running	6 (0.9)
Cycling	4 (0.6)
Hiking/Backpacking	1 (0.2)
Weightlifting	7 (1.1)
Other (e.g., yoga, walking, exercise classes, dance, body surfing)	18 (2.7)

<sup>a</sup> The sum does not equal 100% because percentages were rounded

**Table 4. Patient – Reported Injury Locations**

Injury Location	Frequency (%) <sup>a</sup>
Head/neck	7 (1.1)
Shoulder/arm	1 (0.2)
Elbow/forearm	1 (0.2)
Wrist/hand	5 (0.8)
Trunk/thoracic spine	5 (0.8)
Low back/pelvis	46 (7.0)
Hip/thigh	67 (10.2)
Knee/leg	36 (5.5)
Ankle/foot	18 (2.7)
Other	6 (0.9)
Not Reported	453 (69.2)

<sup>a</sup> The sum does not equal 100% because percentages were rounded

**SCALE STRUCTURE OF THE HOOS SCALE**

The proposed CFA model of the HOOS did not meet contemporary fit recommendations (CFI = 0.847; TLI = 0.836; IFI = 0.847; RMSEA = 0.098; [Figure 1](#)). Correlations between the first-order latent constructs (e.g., symptoms and pain) were high (0.80-0.96; [Figure 1](#)). Modification indices indicated a number of meaningful cross-loadings between several items (e.g., item 6 and item 37 [134.58]) were present.<sup>26</sup> Additionally, modification indices revealed the incorporation of error correlations amongst several items (e.g., item 24 [putting on socks/stockings] and 26 [taking off socks/stockings; 344.25]) would improve model fit (CFI = 0.934; TLI = 0.927; IFI = 0.934; RMSEA = 0.065).

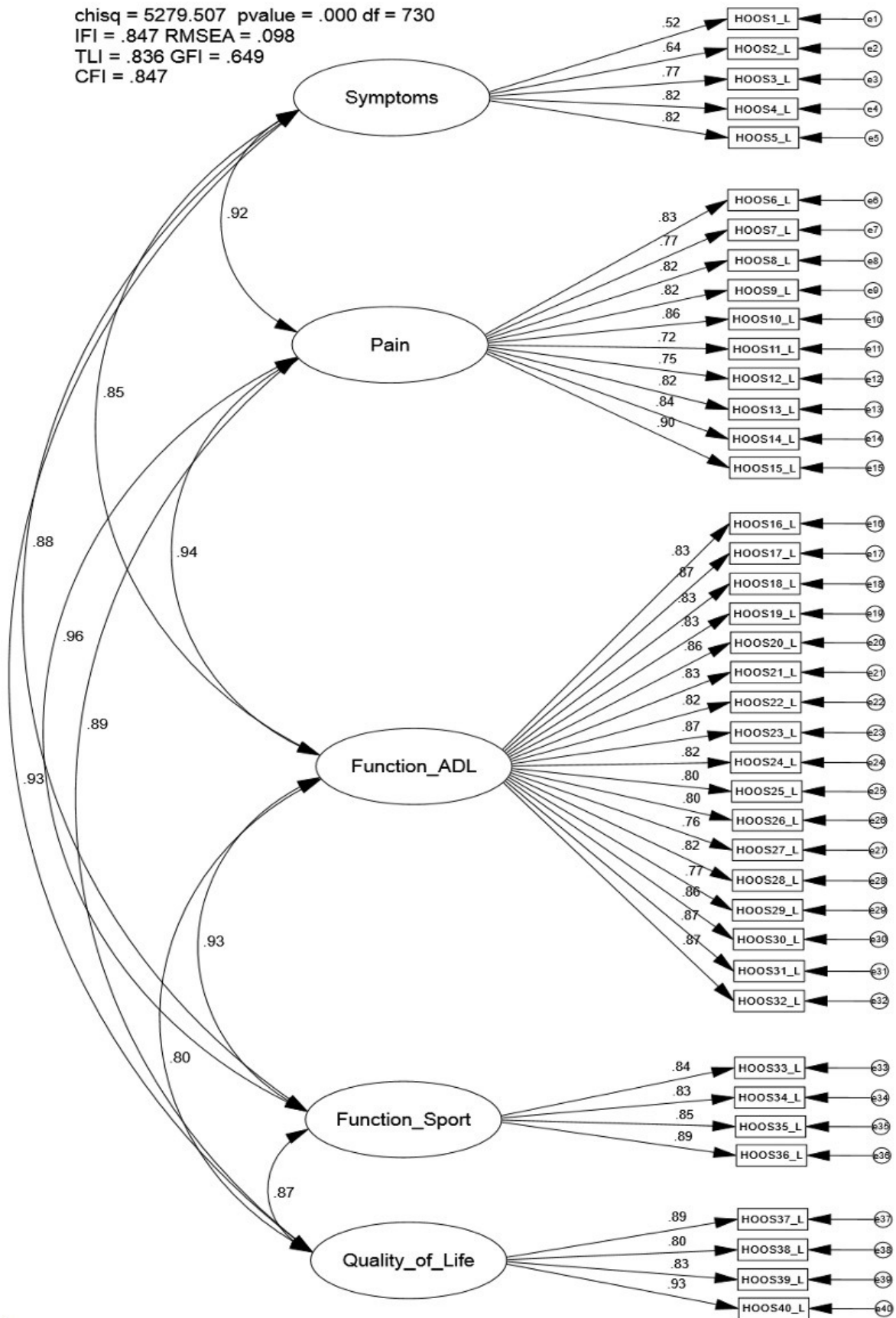
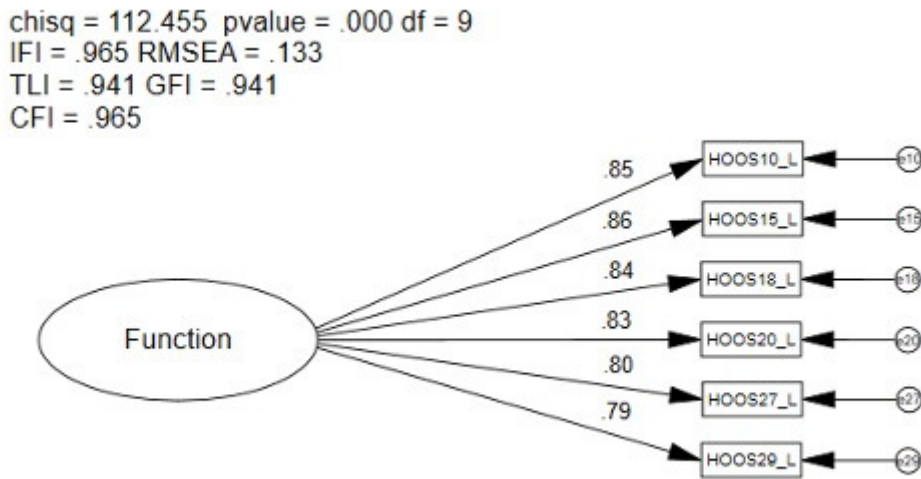


Figure 1. The Hip Dysfunction and Osteoarthritis Score (HOOS) scale hierarchical confirmatory factor analysis measurement model with standardized loadings (n = 656).



**Figure 2. The HOOS-JR scale hierarchical confirmatory factor analysis measurement model with standardized loadings (n = 656).**

SCALE STRUCTURE OF THE HOOS-JR SCALE

The CFA model fit indices of the HOOS-JR met some, but not all contemporary recommendations (CFI = 0.965; TLI = 0.941; IFI = 0.965; RMSEA = 0.133; Figure 2); however, CFI and IFI values exceeded recommendations and loadings were statistically significant (p < 0.001; Figure 2). Modification indices revealed the incorporation of error correlations between two items (i.e., item 10 [going up or down stairs] and 15 [walking on an uneven surface; 69.57] would improve model fit (CFI = 0.994; TLI = 0.988; IFI = 0.994; RMSEA = 0.059).

SCALE STRUCTURE OF THE HOOS-PS SCALE

The CFA model fit indices of the HOOS-PS met some, but not all, contemporary recommendations (CFI = 0.967; TLI = 0.933; IFI = 0.967; RMSEA = 0.137; Figure 3); loadings were statistically significant (p < 0.001; Figure 3). Modification indices revealed that the incorporation of error correlations between two items (i.e., item 16 [descending stairs] and 15 [getting in and out of the bath; 28.90] would improve model fit (CFI = 0.986; TLI = 0.966; IFI = 0.986; RMSEA = 0.098).

SCALE STRUCTURE OF THE HOOS-12 SCALE

The CFA model fit indices of the HOOS-12 did not meet contemporary recommendations (CFI = 0.906; TLI = 0.878; IFI = 0.906; RMSEA = 0.147; Figure 4). Correlations between the first-order latent constructs (e.g., pain and function) were particularly high (0.84-0.98) (Figure 4). Modification indices indicated meaningful cross-loadings between several items (e.g., item 6 and item 37; 57.77) and constructs (e.g., error 37 and pain; 43.34).<sup>26</sup> Additionally, modification indices revealed the incorporation of error correlations between several items (i.e., item 6 [how often is your hip pain] and 37 [how often are you aware of your hip problem;

218.42]) would improve model fit (CFI = 0.976; TLI = 0.965; IFI = 0.976; RMSEA = 0.079).

MULTIGROUP INVARIANCE TESTING ACROSS INJURY SUBGROUP FOR THE HOOS-JR AND HOOS-PS

Given that several model fit indices met recommended fit criteria for the HOOS-JR (i.e., CFI, IFI) and the HOOS-PS (i.e., CFI, TLI), multigroup invariance testing was performed across two subgroups: a hip pathology group (i.e., those diagnosed with hip OA and/or those who underwent a THA) and a non-hip pathology group. The analyses were conducted on the hip pathology group (n = 48; CFI = 0.940), along with a random sample of those who had no self-reported hip pathology (n = 94; CFI = 0.954).

For the HOOS-JR, the initial model (configural) met some recommended model fit indices (CFI = 0.95;  $\chi^2 = 48.01$ ; TLI = 0.914; IFI = 0.95; RMSEA = 0.067; Table 5), indicating equal form between groups on the one factor, 6-item model. The metric model (i.e., equal loadings) passed both the CFI<sub>DIFF</sub> and the  $\chi^2_{DIFF}$  tests, which warranted testing of equal latent variances. After constraining the variances to be equal, the metric model did not pass the CFI<sub>DIFF</sub> or the  $\chi^2_{DIFF}$  test, indicating variances were not equal between groups. When variances were not constrained to be equal, the hip pathology group reported scores with more variance compared to the group without a hip pathology. The scalar model (i.e., equal loadings and intercepts) also passed both the CFI<sub>DIFF</sub> and the  $\chi^2_{DIFF}$  tests (Table 5), which warranted assessment of the equal means. When the means were constrained to be equal, the model did not pass the CFI<sub>DIFF</sub> or the  $\chi^2_{DIFF}$  tests (Table 5), which indicated differences in means between scores. When means were not constrained to be equal, the hip pathology group reported higher mean scores (i.e., more hip dysfunction) than the group without a hip pathology.

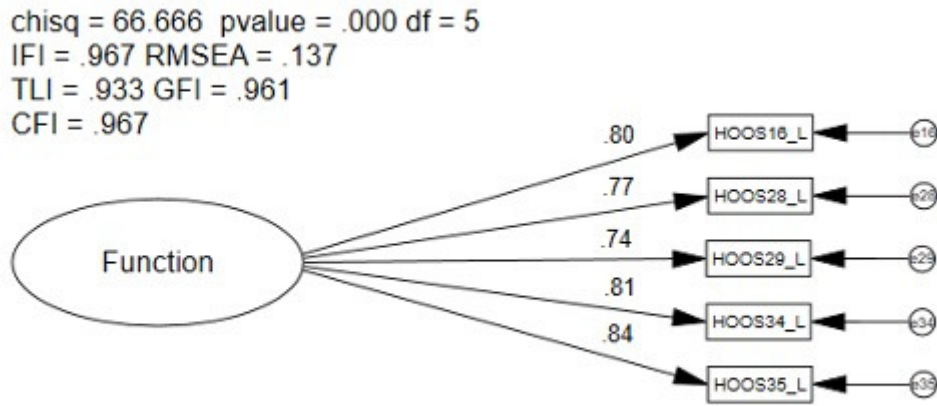


Figure 3. The HOOS-PS hierarchical confirmatory factor analysis measurement model with standardized loadings (n = 656).

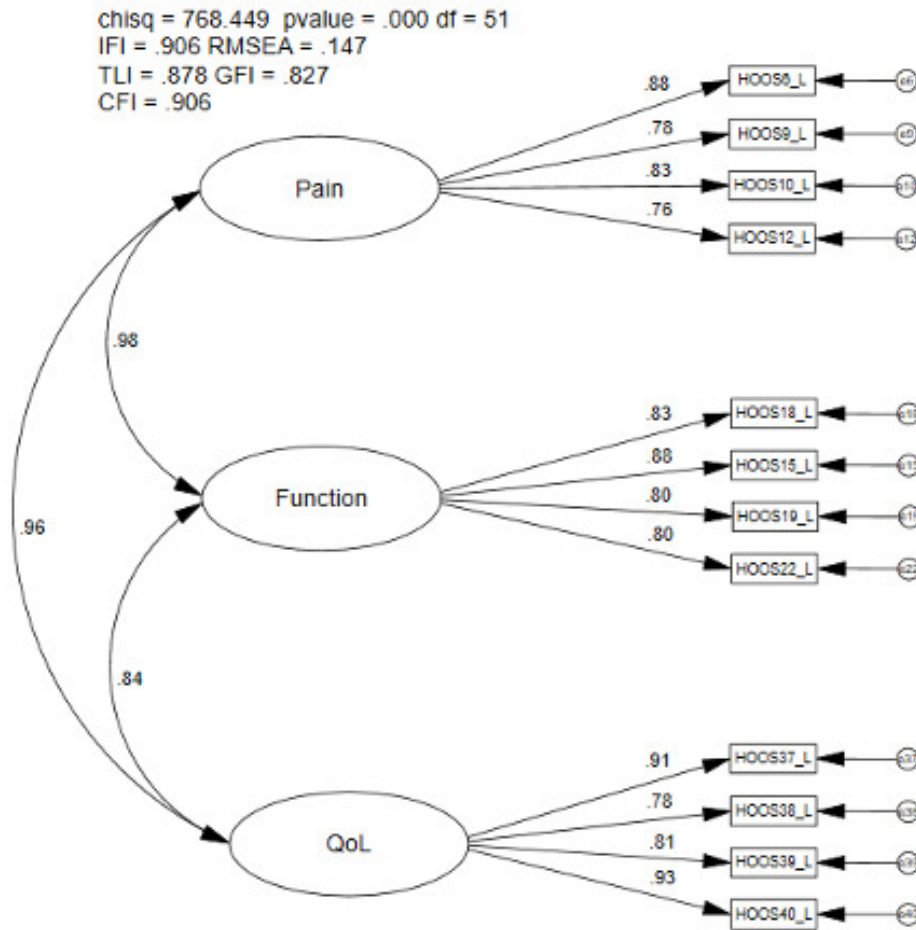


Figure 4. The HOOS-12 item scale hierarchical confirmatory factor analysis measurement model with standardized loadings (n = 656).

For the HOOS-PS, the initial model (configural) met some, but not all, model fit indices (CFI = 0.948;  $\chi^2 = 48.01$ ; TLI = 0.88; IFI = 0.942 RMSEA = 0.13; Table 6). The metric model (i.e., equal loadings) did not pass the CFI<sub>DIFF</sub> or the

$\chi^2_{DIFF}$  tests, which indicated that the meaning of the items was not the same across groups. As such, further exploration of the multigroup invariance testing procedures was not warranted on the HOOS-PS.



**Table 5. Goodness-of-fit indices for Multi-Group Invariance across Hip Pathology**

HOOS-JR	$\chi^2$	df	$\chi^2_{diff}$ (df <sub>diff</sub> )	CFI	CFI <sub>diff</sub>	TLI	RMSEA
OA/THR (n = 48)	23.256	9	----	0.94	----	0.9	0.184
No Hip Pathology (n = 92)	24.684	9	----	0.954	----	0.924	0.138
Configural (equal form)	48.01	18	----	0.948	----	0.914	0.11
Metric (equal loadings)	59.317	23	11.307 (5)	0.938	0.01	0.919	0.107
Equal factor variances*	66.9	24	<b>18.89 (6)</b>	0.926	<b>0.022</b>	0.908	0.114
Scalar (equal indicator intercepts)	65.079	28	17.069 (10)	0.936	0.01	0.932	0.098
Equal latent means*	108.187	29	<b>60.177 (11)</b>	0.864	<b>0.084</b>	0.859	0.141

\* = Substantive questions; Bolded = did not meet cuff off criteria

**Table 6. Goodness-of-fit indices for Multi-Group Invariance across Hip Pathology**

HOOS - PS	$\chi^2$	df	$\chi^2_{diff}$ (df <sub>diff</sub> )	CFI	CFI <sub>diff</sub>	TLI	RMSEA
OA/THR (n = 48)	15.06	5	----	0.94	----	0.87	0.21
No Hip Pathology (n = 92)	18.14	5	----	0.94	----	0.89	0.17
Configural (equal form)	33.24	10	----	0.94	----	0.88	0.13
Metric (equal loadings)	45.32	14	12.08 (4)	0.919	<b>0.02</b>	0.885	0.13
Equal factor variances*	62.50	15	<b>29.26 (5)</b>	0.877	<b>0.06</b>	0.837	0.15
Scalar (equal indicator intercepts)	50.86	18	17.62 (8)	0.915	<b>0.025</b>	0.906	0.12
Equal latent means*	94.95	19	<b>61.71 (9)</b>	0.804	<b>0.14</b>	0.795	0.17

\* = Substantive questions; Bolded = did not meet cuff off criteria

#### MULTIGROUP INVARIANCE TESTING ACROSS ACTIVITY LEVEL SUBGROUPS FOR THE HOOS-JR AND THE HOOS-PS

For the HOOS-JR, the initial model (configural) met some recommended model fit indices (0.95;  $\chi^2 = 171.91$ ; TLI = 0.92; IFI = 0.95; RMSEA = .076; Table 7), which indicated equal form between groups on the one factor, 6-item model. The metric model (i.e., equal loadings) passed both the CFI<sub>DIFF</sub> and the  $\chi^2_{DIFF}$  tests, which warranted testing of equal latent variances. After constraining the variances to be equal, the metric model did not pass the CFI<sub>DIFF</sub> or the  $\chi^2_{DIFF}$  test which indicated variances were not equal between groups. When variances were not constrained to be equal, the inactive group reported scores with more variance compared to the other three groups. The scalar model (i.e., equal loadings and intercepts) also passed both the CFI<sub>DIFF</sub> and the  $\chi^2_{DIFF}$  tests (Table 7), which warranted assessment of the equal means. When the means were constrained to be equal, the model did not pass the CFI<sub>DIFF</sub> or the  $\chi^2_{DIFF}$  tests (Table 7), which indicated differences in means between scores. When means were not constrained to be equal, the inactive group reported higher mean scores (i.e., more hip dysfunction) than the active groups.

For the HOOS-PS, the initial model (configural) met model fit indices (CFI = 0.97;  $\chi^2 = 79.03$ ; TLI = 0.93; IFI = 0.97; RMSEA = 0.067; Table 8), which indicated equal form of the one factor, 5-item model between groups. The metric model (i.e., equal loadings) passed both the CFI<sub>DIFF</sub> and the  $\chi^2_{DIFF}$  test, which warranted testing of equal latent variances. After constraining the variances to be equal, the

model did not pass the CFI<sub>DIFF</sub> or the  $\chi^2_{DIFF}$  test, which indicated differences in variance between groups. When variances were not constrained to be equal, the inactive group reported more variance in scores than the active groups (i.e., low-, moderate-, and high activity). The scalar model (i.e., equal loadings and intercepts) also passed both the CFI<sub>DIFF</sub> and the  $\chi^2_{DIFF}$  tests (Table 8), which warranted assessment of the equal means model. When means were constrained to be equal, the model did not pass the CFI<sub>DIFF</sub> or the  $\chi^2_{DIFF}$  tests (Table 8); when means were not constrained to be equal, individuals in the inactive group reported higher mean scores (i.e., more hip dysfunction) than the active groups.

#### DISCUSSION

The purpose of this study was to examine the psychometric properties of the published 40-item HOOS, HOOS-PS, HOOS-JR, and HOOS-12 scales by using contemporary CFA and multigroup invariance testing procedures in a larger and more diverse physically active sample. Confirmatory factor analysis procedures were used as an approach to examine these scales for use in clinical practice and research, while invariance testing procedures helped assess for item-level bias and substantive differences between groups.<sup>27</sup> Previous literature demonstrated good model fit of individual constructs (i.e., pain, function, and QoL)<sup>17</sup>; however, previous researchers failed to provide model fit of the full latent variable scale model as recommended<sup>26,31</sup> to assess scale properties for use in practice and research.<sup>26,31,33</sup> The

**Table 7. Goodness-of-fit indices for Multi-Group Invariance across Physical Activity**

HOOS-JR	$\chi^2$	df	$\chi^2_{diff}$ (df <sub>diff</sub> )	CFI	CFI <sub>diff</sub>	TLI	RMSEA
Inactive (n = 40)	17.56	9	----	0.962	----	0.937	0.156
Low Activity (n = 225)	61.23	9	----	0.953	----	0.921	0.161
Moderate Activity (n = 276)	35.79	9	----	0.974	----	0.957	0.104
High Activity (n = 113)	57.03	9	----	0.878	----	0.797	0.218
Configural (equal form)	171.91	36	----	0.951	----	0.919	0.076
Metric (equal loadings)	198.63	51	26.72 (15)	0.947	0.004	0.938	0.067
Equal factor variances*	227.46	54	<b>55.55 (18)</b>	0.938	<b>0.013</b>	0.931	0.070
Scalar (equal indicator intercepts)	219.70	66	47.79 (30)	0.945	0.006	0.950	0.060
Equal latent means*	244.52	69	<b>72.61 (33)</b>	0.937	<b>0.014</b>	0.945	0.063

\* = Substantive questions; Bolded = did not meet cuff off criteria

**Table 8. Goodness-of-fit indices for Multi-Group Invariance across Physical Activity**

HOOS - PS	$\chi^2$	df	$\chi^2_{diff}$ (df <sub>diff</sub> )	CFI	CFI <sub>diff</sub>	TLI	RMSEA
Inactive (n = 40)	11.31	5	----	0.966	----	0.932	0.18
Low Activity (n = 225)	26.13	5	----	0.972	----	0.943	0.137
Moderate Activity (n = 276)	16.57	5	----	0.979	----	0.957	0.099
High Activity (n = 113)	22.84	5	----	0.899	----	0.799	0.178
Configural (equal form)	79.03	20	----	0.966	----	0.932	0.067
Metric (equal loadings)	99.79	32	20.76 (12)	0.961	0.005	0.951	0.057
Equal factor variances*	142.90	35	<b>63.869 (15)</b>	0.938	<b>0.028</b>	0.929	0.069
Scalar (equal indicator intercepts)	121.12	44	42.09 (22)	0.956	0.005	0.96	0.052
Equal latent means*	140.43	47	<b>61.4 (17)</b>	0.946	<b>0.02</b>	0.954	0.055

\* = Substantive questions; Bolded = did not meet cuff off criteria

current results indicate the original HOOS and HOOS-12 do not meet recommended measurement criteria for this sample of physically active participants. Therefore, caution is warranted if using results from either measure for research or clinical practice. The HOOS-JR and the HOOS-PS demonstrated stronger evidence supporting their use given the CFA and multigroup invariance findings. Further exploration to determine when to use the scales and when the measurement properties may not be sufficient for assessing group differences in larger samples of physically active patients with and without hip pathology is warranted to confirm or refute our findings.

**CONFIRMATORY FACTOR ANALYSIS OF THE ORIGINAL 40-ITEM HOOS SCALE**

The original five factor, 40-item HOOS scale structure was not supported in our study.<sup>26,33</sup> Poor model fit indices, along with high correlation values between latent constructs, indicates potential multicollinearity and a lack of unique constructs. Additionally, the modification indices revealed that model fit could be substantially improved if numerous modifications in the model (e.g., error-terms were correlated) were instituted.<sup>26,33</sup> Assessment of the error-term cross-loadings revealed that most of the items shared commonalities.<sup>26,35</sup> Correlation of the error terms

may indicate the presence of overlapping items, or items that are perceived to ask similar questions.<sup>26</sup> Further, there were concerns with Cronbach's alpha values; in our sample, the high values (0.84-0.98) were similar to previously reported levels (0.75-0.98)<sup>13,18,23</sup> and may be indicative of potential item redundancy.<sup>20,36</sup> The high correlation values between constructs and items, along with high Cronbach's alpha values, re-affirmed multicollinearity as a concern and may indicate respondents are unable to differentiate between the items used to measure different constructs.<sup>20,36</sup>

The current findings suggest the model may be improved by re-writing items or by removing items from the original model.<sup>26,31</sup> Furthermore, the results make it difficult to conclude that items in the constructs are measuring unique phenomena.<sup>26,31,33</sup> Also, exploration may be warranted to determine if the correlated errors are theoretically justified and to determine when the inclusion of those correlations are warranted in research.<sup>26,35</sup> Thus, the instrument may be improved through exploratory procedures (i.e., EFA procedures) to help determine if a more concise instrument can be identified from the originally developed item.<sup>20,26,31</sup> However, given the design of the HOOS items, further modification may be necessary. For example, many HOOS items are double-barreled questions (i.e., asking more than one question in an item), which may result in analysis complications because the respondent may not know which as-

pect of the item to respond to for their scored response and may cause confusion and generate inconsistent results.<sup>37</sup> As such, it would be prudent for researchers to rewrite items or provide fewer overlapping examples which may result in improved model fit and more precise assessment of the patient experience.<sup>37</sup>

The current CFA analysis approach on the full HOOS scale provides insight where previous studies separated the dimensions to conduct CFAs on individual dimensions of the scale (i.e., the development of the HOOS-12).<sup>17</sup> To the authors knowledge, this study is the first to perform CFA procedures on the full HOOS reflective latent variable model. Analyses examining psychometric properties of a scale should first examine the full model prior to conducting exploratory procedures and item removal.<sup>26,31,33</sup> The model fit indices found in our study were substantially lower than those previously reported for the pain and function dimensions when the constructs were examined individually.<sup>17</sup> These results demonstrate the importance of testing the full model before recommending a multi-dimensional scale for use in practice or research.<sup>30,31</sup>

#### CONFIRMATORY FACTOR ANALYSIS OF THE HOOS-JR

The HOOS-JR met recommended CFI and IFI values<sup>26,33</sup>; model fit concerns (e.g., high item cross-loadings) may have contributed to a reduced overall model fit. Additionally, the modification indices revealed that model fit could be substantially improved if modifications in the model were instituted.<sup>26,33</sup> Assessment of the error term cross-loadings identified in the modification indices indicated two items (i.e., item 10 [going up or down stairs] and item 15 [walking on an uneven surface]) shared commonalities, which may warrant further exploration as including the error term covariances may be appropriate in certain analysis situations.<sup>26,35</sup> Similar to the HOOS, another potential explanation for poor model fit indices could be item design. It may be beneficial to address double-barreled questions and overlapping items to improve model fit and reduce response burden for respondents.<sup>15,37</sup>

The HOOS-JR was subjected to multigroup invariance testing by injury type and activity level as certain model fit criteria (i.e., CFI, IFI) thresholds were met. The multigroup invariance findings across injury type provide some evidence for scale validity. As the model met criteria for measurement invariance, group differences for variances and latent means could be assessed as these differences could be considered true differences as opposed to differences due to item bias or measurement error.<sup>26,31</sup> If the HOOS-JR was valid scale, it would be expected that respondents who had hip OA or previous history of a THA would report higher mean scores with greater variances if the scale is measuring the intended phenomenon. A higher score on the HOOS-JR construct indicates those respondents have more difficulty related to pain and function,<sup>16,17</sup> while a finding of more score variance and higher mean scores for impaired function and pain in the injured group would be expected because hip OA is one of the leading causes of decreased function due to pain.<sup>10</sup> Our results indicate the individuals with hip pathology reported larger amounts

of variance and higher mean scores compared to the no hip pathology group. These substantive findings provide support that the HOOS-JR is capturing valid group differences among those who are suffering from a hip injury/dysfunction and those who are not. Thus, clinicians and researchers could assess score differences between these groups on the HOOS-JR.

Multigroup invariance testing was then performed by activity level subgroups. Group differences for variances and means were also found between activity level subgroups, which also support the validity of the HOOS-JR. A higher score on the HOOS-JR constructs indicates those respondents have greater difficulty related to pain and function<sup>16,17</sup> and it could be theorized that individuals with higher levels of hip dysfunction (e.g., pain) would be less active than those with lower levels of dysfunction. The current findings reveal individuals who were classified as inactive reported larger amounts of variance in their scores and exhibited higher mean scores (i.e., more pain and decreased function) compared to those who were more active (i.e., low-, moderate-, and high-activity). The findings indicate the group differences likely represent true score differences as opposed to measurement error; thus, our results provide substantive support for scale validity, given that the HOOS-JR identified higher dysfunction in inactive patients who likely alter activity levels due to hip pain and dysfunction. Further analysis of the inactive group supports this theory as 42% (N = 17) of these participants reported a current physical injury, and 20% (N = 8) reported a previous injury to their hip. Thus, the findings support the HOOS-JR is capturing valid group differences in those who are less active and suffering from a hip injury/dysfunction as compared to healthy, active respondents.

#### CONFIRMATORY FACTOR ANALYSIS OF THE HOOS-PS SCALE

The one factor, 5-item HOOS-PS met the recommended levels for CFI, GFI, and TLI<sup>26,33</sup>; however, additional model fit concerns such as item cross-loadings may have contributed to reduced overall model fit. Like the HOOS and the HOOS-JR, a potential explanation for poor model fit indices could be item design. Assessment of cross-loadings identified through the modification indices was performed; review of the items did not indicate a theoretical justification for the shared commonalities.<sup>26,35</sup> Thus, sound rationale for further exploration of the correlation of error terms was not identified.<sup>26,35</sup> However, it may be beneficial to address double-barreled questions and overlapping items to improve model fit.<sup>15,37</sup>

Next, the HOOS-PS was subjected to multigroup invariance testing by injury type. As model fit indices for the metric invariance model were not met,<sup>26</sup> the use of this scale may not be appropriate for examining group differences and differences in scores between respondents who have a hip pathology and those who do not in its current form. Without meeting multigroup invariance testing recommendations, it should not be assumed that score differences between healthy or injured respondents are true differences and not measurement error.<sup>26</sup> Multigroup invariance test-

ing should be performed again in a larger sample of healthy and hip injured respondents to confirm or refute our findings.

Lastly, when performing invariance testing by activity level subgroups, evidence was found to support scale structure with the configural, metric, and scalar results.<sup>26</sup> Group differences in variances and means for function were found between activity levels. Individuals who were classified as being inactive had more variance in their responses compared to those who were classified as active. A higher score on the HOOS-PS constructs indicates those respondents have more difficulty related to physical function.<sup>10,38</sup> A finding of more score variance and higher mean scores in physical function of the inactive group would be expected, as some participants in this sample were more likely to have difficulties pertaining to their hip while performing physical activity. Thus, the substantive findings provide support that the HOOS-PS is capturing valid group differences between activity levels in our sample, which provides theoretical support for the HOOS-PS.

#### CONFIRMATORY FACTOR ANALYSIS OF THE HOOS-12 SCALE

As the HOOS-12 model did not meet the recommended model fit indices in our sample<sup>26,33</sup>; a number of concerns regarding model fit were present. First, a high correlation values between latent constructs indicate potential multicollinearity and an inability of the items to measure unique constructs. The modification indices also revealed model fit could be substantially improved if modifications in the model (e.g., error terms were correlated) were made.<sup>26,33</sup> Additionally, concerns with Cronbach's alpha values were present; the values were high (0.88-0.91), which are similar to previously reported levels (0.77-0.95)<sup>15</sup> indicates potential item redundancy.<sup>20,36</sup> Lastly, assessment of cross-loadings identified through the modification indices revealed that some, but not all, of the items shared commonalities which could be justified.<sup>26,35</sup> As such, further exploration may be warranted to determine if and when the correlated errors should be included in a model.<sup>26,35</sup> High correlation values between constructs and items, coupled with high Cronbach's alpha values, reiterate the concern regarding the presence of multicollinearity bordering on singularity and the potential need to remove items, alter items to improve clarity, or to develop new items which better measure the intended construct.<sup>26,31</sup>

#### LIMITATIONS AND FUTURE RESEARCH

While the present study identified concerns regarding the factorial validity in the HOOS, the HOOS-JR, the HOOS-PS and the HOOS-12, there are still limitations to consider. The current sample was larger than those used for most studies on the HOOS, but this sample was comprised mostly of self-reported healthy individuals. Moreover, the sample used included few participants who had been diagnosed with hip OA, THA, or injury to their hip. Of note,

a sub-sample of the healthy participants was used in the multigroup invariance testing due to the limited sample of injured participants. Utilizing such a small sample size may impact the statistical power of the test and result in model misspecification, which is why larger sample sizes are recommended.<sup>26</sup> As such, further exploration is warranted using larger, more evenly distributed and diverse samples. Also, due to the limited sample size and the current clinical application of the scale, invariance testing using the HOOS-JR was limited to the originally proposed model; however, future research should explore the validity of including the error term correlation identified and how it influences findings.

The sample of participants also responded to all 40-items of the HOOS. Thus, it is possible that responses to the short forms were influenced by the additional items not on the scale. Therefore, future research should be completed on a sample of participants who only responded to the items on the instrument. The authors also did not conduct long-term follow-up nor compare the results of the modified scale with another criterion scale. Due to study design (i.e., collection at one time point), we could not perform test-retest reliability, assess the minimal detectable change, assess responsiveness (e.g., the minimal clinically important differences [MCIDs]), or perform longitudinal invariance testing. Future research should aim to assess longitudinal invariance and measures of instrument precision (e.g., MCIDs) to fully establish the psychometric properties of each scale and to provide guidance for use of the scales in clinical practice and research.

#### CONCLUSIONS

In conclusion, the scale structure of the original HOOS and HOOS-12 were not supported in the current study. Analyses found preliminary evidence to support the use of the HOOS-JR and HOOS-PS as psychometrically sound instruments and multigroup invariance testing results provided substantive support for these scales measuring the intended phenomenon and ability to assess true group differences in certain situations. Clinicians and researchers who utilize the scales, should do so with caution regarding their limitations or untested properties (e.g., longitudinal invariance testing). Thus, more research is warranted to establish the full psychometric properties of these scales and identify an improved version which meets contemporary recommendations to measure the multi-dimensional experience of patient disability following hip pathology.

#### CONFLICTS OF INTEREST

The authors report no conflicts of interest

Submitted: May 12, 2022 CST, Accepted: November 03, 2022 CST



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc/4.0> and legal code at <https://creativecommons.org/licenses/by-nc/4.0/legalcode> for more information.

## REFERENCES

1. Ruysen-Witrand A, Fernandez-Lopez C, Gossec L, Anract P, Courpied J, Dougados M. Psychometric properties of the OARSI/OMERACT osteoarthritis pain and functional impairment scales: ICOAP, KOOS-PS and HOOS-PS. *Clin Exp Rheum-Incl Supplements*. 2011;29(2):231.
2. Botha-Scheepers S, Watt I, Rosendaal FR, Breedveld FC, Hellio le Graverand MP, Kloppenburg M. Changes in outcome measures for impairment, activity limitation, and participation restriction over two years in osteoarthritis of the lower extremities. *Arthrit Care Res*. 2008;59(12):1750-1755. doi:10.1002/art.24080
3. Jette AM. Outcomes research: shifting the dominant research paradigm in physical therapy. *Phys Ther*. 1995;75(11):965-970. doi:10.1093/ptj/75.11.965
4. Vela LI, Denegar CR. The disablement in the physically active scale, part II: The psychometric properties of an outcomes scale for musculoskeletal injuries. *J Athl Train*. 2010;45(6):630-641. doi:10.4085/1062-6050-45.6.630
5. Vingård E, Alfredsson L, Malchau H. Osteoarthrosis of the hip in women and its relationship to physical load from sports activities. *Am J Sports Med*. 1998;26(1):78-82. doi:10.1177/03635465980260013101
6. Felson DT, Zhang Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. *Arthritis Rheum*. 1998;41(8):1343-1355.
7. Nilsson AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS) – validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord*. 2003;4(1):10. doi:10.1186/1471-2474-4-10
8. Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G. A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. *Arthrit Care Res*. 1999;12(5):331-335.
9. Davis AM, Badley EM, Beaton DE, et al. Rasch analysis of the western ontariomcmaster (WOMAC) osteoarthritis index: results from community and arthroplasty samples. *J Clin Epidemiol*. 2003;56(11):1076-1083. doi:10.1016/s0895-4356(03)0179-3
10. Davis AM, Perruccio AV, Canizares M, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. *Osteoarthr Cartil Open*. 2008;16(5):551-559. doi:10.1016/j.joca.2007.12.016
11. Sun Y, Sturmer T, Gunther KP, Brenner H. Reliability and validity of clinical outcome measurements of osteoarthritis of the hip and knee – a review of the literature. *Int J Rheum Dis*. 1997;16(2):185-198. doi:10.1007/bf02247849
12. Nilsson A, Bremander A. Measures of hip function and symptoms: Harris hip score (HHS), hip disability and osteoarthritis outcome score (HOOS), Oxford hip score (OHS), Lequesne index of severity for osteoarthritis of the hip (LISOH), and American Academy of orthopedic surgeons (AAOS) hip and knee questionnaire. *Arthrit Care Res*. 2011;63(S11):S200-S207. doi:10.1002/acr.20549
13. Ornetti P, Parratte S, Gossec L, et al. Cross-cultural adaptation and validation of the French version of the Hip disability and Osteoarthritis Outcome Score (HOOS) in hip osteoarthritis patients. *Osteoarthr Cartil Open*. 2010;18(4):522-529. doi:10.1016/j.joca.2009.12.007
14. Nilsson A, Petersson I, Roos EM, Lohmander L. Predictors of patient relevant outcome after total hip replacement for osteoarthritis: a prospective study. *Ann Rheum Dis*. 2003;62(10):923-930. doi:10.1136/ard.62.10.923
15. Gandek B, Roos EM, Franklin PD, Ware JE Jr. A 12-item short form of the Hip disability and Osteoarthritis Outcome Score (HOOS-12): tests of reliability, validity and responsiveness. *Osteoarthr Cartil Open*. 2019;27(5):754-761. doi:10.1016/j.joca.2018.09.017
16. Lyman S, Lee YY, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: a short-form hip replacement survey. *Clin Orthop Relat Res*. 2016;474(6):1472-1482. doi:10.1007/s11999-016-4718-2
17. Gandek B, Roos EM, Franklin PD, Ware JE Jr. Item selection for 12-Item short forms of the knee injury and osteoarthritis outcome score (KOOS-12) and hip disability and osteoarthritis outcome score (HOOS-12). *Osteoarthr Cartil Open*. 2019;27(5):746-753. doi:10.1016/j.joca.2018.11.011

18. De Groot IB, Reijman M, Terwee CB, et al. Validation of the Dutch version of the Hip disability and Osteoarthritis Outcome Score. *Osteoarthr Cartil Open*. 2007;15(1):104-109. [doi:10.1016/j.joca.2006.06.014](https://doi.org/10.1016/j.joca.2006.06.014)
19. Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences*. Vol 663. Houghton Mifflin College Division; 2003.
20. Leech NL, Barrett KC, Morgan GA. *IBM SPSS for Intermediate Statistics: Use and Interpretation*. Routledge; 2014.
21. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess*. 2003;80(1):99-103. [doi:10.1207/s15327752jpa8001\\_18](https://doi.org/10.1207/s15327752jpa8001_18)
22. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ*. 2018;48(6):1273-1296. [doi:10.1007/s11165-016-9602-2](https://doi.org/10.1007/s11165-016-9602-2)
23. Klässbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score: An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. *Scand J Rheumatol*. 2003;32(1):46-51. [doi:10.1080/03009740310000409](https://doi.org/10.1080/03009740310000409)
24. Panayides P. Coefficient alpha: interpret with caution. *Eur J Psychol*. 2013;9(4):687-696. [doi:10.5964/ejop.v9i4.653](https://doi.org/10.5964/ejop.v9i4.653)
25. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vision Sci*. 2007;84(8):663-674. [doi:10.1097/OPX.0b013e318141fe75](https://doi.org/10.1097/OPX.0b013e318141fe75)
26. Kline RB. *Principles and Practice of Structural Equation Modeling*. Guilford publications; 2015.
27. Bryant FB, Yarnold PR. Principal-components analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold PR, eds. *Reading and Understanding Multivariate Statistics*. American Psychological Association; 1995:99-136.
28. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. *J Med Internet Res*. 2017;19(8):e290. [doi:10.2196/jmir.7071](https://doi.org/10.2196/jmir.7071)
29. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 4th ed. Allyn & Bacon; 2001.
30. Grimm LG, Yarnold PR. *Reading and Understanding Multivariate Statistics*. American Psychological Association; 1995.
31. Brown TA. *Confirmatory Factor Analysis for Applied Research*. Guilford publications; 2015.
32. Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming (multivariate applications series). *New York: Taylor & Francis Group*. 2010;396(1):7384.
33. Byrne BM. *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming*. Routledge; 2013. [doi:10.4324/9780203807644](https://doi.org/10.4324/9780203807644)
34. Aguinis H, Gottfredson RK, Joo H. Best-practice recommendations for defining, identifying, and handling outliers. *Organ Res Methods*. 2013;16(2):270-301. [doi:10.1177/1094428112470848](https://doi.org/10.1177/1094428112470848)
35. Nachtigall C, Kroehne U, Funke F, Steyer R. Pros and cons of structural equation modeling. *Methods Psychological Research Online*. 2003;8(2):1-22.
36. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clin Ther*. 1996;18(5):979-992. [doi:10.1016/s0149-2918\(96\)80054-3](https://doi.org/10.1016/s0149-2918(96)80054-3)
37. Dillman DA, Smyth JD, Christian LM. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons; 2014.
38. Davis AM, Perruccio AV, Canizares M, et al. Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. *Osteoarthr Cartil Open*. 2009;17(7):843-847. [doi:10.1016/j.joca.2009.01.005](https://doi.org/10.1016/j.joca.2009.01.005)