

A Systematic Review of Measurement Properties of the Forgotten Joint Score (FJS) Based on the COSMIN Guideline

Rui Guo¹, Jiachen Liang², Han Li³, Qing Sun^{3,*}, Hanfeng Wang¹

¹Shaanxi University of Chinese Medicine, Xianyang 712046, Shaanxi, China

²Tuina Department, First Teaching Hospital of Tianjin University of Traditional Chinese Medicine
National Clinical Research Center for Chinese Medicine Acupuncture and Moxibustion, Tianjin 300193, China

³Xi'an Hospital of Traditional Chinese Medicine, Xi'an 710021, Shaanxi, China

*Correspondence Author

Abstract: ***Objective:** This study aimed to comprehensively assess the measurement properties of the Forgotten Joint Score (FJS) in patients undergoing total knee arthroplasty (TKA). **Methods:** A systematic review of the measurement properties of the FJS was conducted according to the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN). We systematically searched PubMed, Web of Science, The Cochrane Library, Medline, Embase, Psych Info, CNKI, and VIP Database from inception to December 1, 2023. Studies focusing on post-TKA patients were included, and the measurement properties of the FJS were systematically described. **Results:** Eleven studies covering 11 language versions of the FJS were included. Methodological quality for content validity was rated as adequate, with an overall measurement property rating of sufficient. For structural validity, methodological quality was doubtful and the measurement property was indeterminate. Both internal consistency and reliability were rated as very good in methodological quality and sufficient as measurement properties. Structural validity, cross-cultural validity, and criterion validity could not be assessed due to insufficient information. Responsiveness showed doubtful methodological quality and insufficient measurement properties. Hypothesis testing was rated as inadequate in methodological quality and insufficient as a measurement property. GRADE ratings indicated sufficient content validity, indeterminate structural validity, high-quality evidence for internal consistency and reliability, and insufficient evidence for measurement error and responsiveness. **Conclusion:** The FJS demonstrates good reliability and validity. Future research should fill the gaps in the evaluation of structural validity, cross-cultural validity, and criterion validity, and improve hypothesis testing studies, so as to fully and accurately determine whether the FJS is suitable for use in Chinese TKA patients.*

Keywords: Total knee arthroplasty, COSMIN guideline, Assessment tool, Measurement properties, Systematic review.

1. Introduction

Total knee arthroplasty (TKA) is an effective treatment for end-stage knee osteoarthritis, significantly improving joint function, relieving pain, and enhancing quality of life. It has become a commonly used joint reconstruction procedure in orthopedic practice [1]. Accurate assessment of postoperative joint function, pain perception, and quality of life is key to developing individualized rehabilitation plans and adjusting clinical treatment strategies. Patient-reported outcome measures (PROMs) serve as core tools for evaluating outcomes after TKA, and the scientific rigor and reliability of their measurement properties directly determine the value of the assessment [2-4]. Commonly used scales for evaluating TKA outcomes in current clinical practice, such as the Oxford Knee Score (OKS), the New Knee Society Scoring System (New-KSS), and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), have been found to have various measurement deficiencies according to the COSMIN guideline, including insufficient validity, large measurement errors, and poor responsiveness. These shortcomings compromise the credibility of the results and may mislead clinical decision-making [5-7]. The Forgotten Joint Score (FJS), introduced in 2012 [8], focuses on the extent to which patients forget their artificial joint in daily life, thereby intuitively reflecting postoperative discomforts such as pain and stiffness [9]. Owing to its novel perspective and ease of administration, the FJS has been translated into multiple languages and applied in postoperative TKA assessment [10,11].

However, existing studies on the reliability and validity of the FJS [6,10,12,13,14] are fragmented, and no comprehensive systematic evaluation of its measurement properties based on the COSMIN guideline has yet been conducted. Quantitative analyses of scale development details and content validity remain insufficient, and research on key attributes such as cross-cultural validity and criterion validity across different language versions is lacking. Whether the FJS can accurately meet the assessment needs of Chinese TKA patients remains unclear.

COSMIN is an internationally recognized standardized system for evaluating health measurement instruments, enabling a comprehensive assessment of the measurement properties of PROMs. This study performed a systematic review of the measurement properties of the FJS based on the COSMIN guideline, aiming to fill the gaps in evaluating structural validity, cross-cultural validity, and other properties, and to clarify the scientific rigor and reliability of the FJS for use in TKA patients, thereby providing evidence-based support for its appropriate application in postoperative TKA assessment in China.

2. Methods

Following the COSMIN guideline for systematic reviews, this study applied its methodology and practical tools to evaluate the measurement properties of the FJS, with the ultimate goal of facilitating its application and refinement [5].

Table 1: Definitions of measurement properties

Measurement property	Definition
Content validity	The degree to which the content of a PROM adequately reflects the construct to be measured.
Structural validity	The degree to which PROM scores are consistent with hypotheses (e.g., internal relationships, relationships with scores of other instruments, or differences between relevant groups) based on the assumption that the PROM validly measures the target construct.
Internal consistency	The interrelatedness among items.
Reliability	The proportion of total variance in measurements that is due to true differences between patients.
Measurement error	Systematic and random error in patient scores that is not attributable to true changes in the construct to be measured.
Cross-cultural validity/Measurement invariance	The degree to which the performance of items on a translated or culturally adapted PROM reflects the performance of items on the original version.
Construct validity	The degree to which PROM scores adequately reflect the dimensionality of the construct to be measured.
Criterion validity	The degree to which PROM scores adequately reflect a 'gold standard'.
Responsiveness	The ability of a PROM to detect change over time in the construct to be measured.
Interpretability	The degree to which qualitative meaning (i.e., clinical or commonly understood connotations) can be assigned to quantitative PROM scores or changes in scores.

2.1 Inclusion and Exclusion Criteria

Inclusion criteria: (1) patients who had undergone TKA; (2) age > 18 years; (3) full articles reporting details of FJS development; (4) reporting at least one measurement property (Table 1[15]), covering content validity, structural validity, internal consistency, test-retest reliability, measurement error, cross-cultural validity/measurement invariance, criterion validity, and responsiveness [16]; (5) the FJS had to be completed by patients themselves.

Exclusion criteria: (1) studies using the scale solely as an outcome measurement tool (e.g., interventional studies); (2) secondary literature such as reviews, systematic reviews, or meta-analyses; (3) articles for which the full text could not be obtained; (4) articles not published in Chinese or English; (5) duplicate publications.

2.2 Search Strategy

Literature searches were conducted from database inception to December 20, 2023, in the following databases: PubMed, Web of Science, The Cochrane Library, Medline, Embase, Psych Info, CNKI, and VIP Database. Chinese search terms included: forgotten joint score, total knee arthroplasty, reliability, validity, responsiveness. English search terms included: forgotten knee score, forgotten knee, forgotten joint, FJS, TKA, total knee arthroplasty, total knee replacement, knee replacement arthroplasties, knee replacement arthroplasty, knee arthroplasty, replacement arthroplasty knee, arthroplasties knee replacement, knee replacement total, replacement total knee, arthroplasty total knee, arthroplasty knee replacement, structural validity, factor analysis, internal consistency, cronbach, reliability, measurement error, hypotheses testing, cross-cultural validity, measurement invariance, cross cultural validity, criterion validity, responsiveness, content validity. Manual searches were also performed to identify additional relevant studies.

2.3 Study Selection and Data Extraction

Literature management was carried out using Zotero. Two reviewers independently extracted full-text data, applied the exclusion criteria, and collected patient characteristics (including sample size, mean age, sex distribution, disease characteristics, and response rate) as well as measurement characteristics (such as FJS version, follow-up interval, study setting, country, language, and mode of questionnaire administration). All available information on measurement

properties (e.g., content validity, structural validity, internal consistency, reliability, measurement error, cross-cultural validity/measurement invariance, criterion validity, and responsiveness) was also extracted. Any discrepancies between the two reviewers were first resolved through discussion; if no agreement could be reached, a third party was consulted for arbitration, and consensus was ultimately achieved through further discussion.

2.4 Synthesis and Grading of Evidence

2.4.1 Assessment of Methodological Quality

Methodological quality was evaluated using the COSMIN Risk of Bias checklist [17], which covers ten modules: PROM development, content validity, structural validity, internal consistency, reliability, measurement error, cross-cultural validity, criterion validity, hypotheses testing, and responsiveness. A four-point rating system was applied — “very good”, “adequate”, “doubtful”, and “inadequate” — following the worst-score-counts principle.

2.4.2 Assessment of Measurement Properties

Measurement properties were rated according to the COSMIN quality criteria [18], including structural validity, internal consistency, reliability, measurement error, hypotheses testing, cross-cultural validity, criterion validity, and responsiveness. A three-point scale was used: sufficient (+), insufficient (-), or indeterminate (?).

2.4.3 Grading of Evidence

The GRADE approach [19] was applied to rate each measurement property as sufficient (+), insufficient (-), indeterminate (?), or inconsistent (\pm). Each measurement property of the scale was initially set as high-quality evidence and then downgraded based on risk of bias, indirectness, inconsistency, and imprecision, resulting in four levels of evidence quality: high, moderate, low, and very low.

To synthesize the results for each measurement property, hypotheses regarding expected correlations between the FJS and comparator instruments were pre-defined (Table 2). Within the included studies, all correlations between the FJS and comparator tools were examined. The percentage of accepted hypotheses and studies was then calculated by pooling the weighted average of correlations. Disagreements regarding the synthesis of results and grading of evidence

were resolved through discussion [20].

Table 2: Pre-defined hypotheses: expected correlations between FJS and comparator instruments

1	Correlations (or changes) with instruments measuring physical function (e.g., WOMAC physical function subscale, FJS, KOOS, OKS, SF-36) should be >0.50 .
2	Correlations (or changes) with instruments measuring pain (e.g., WOMAC, OKS, or KOOS pain subscale) or stiffness (e.g., WOMAC stiffness subscale) should be $0.30-0.50$.
3	Correlations (or changes) with instruments measuring unrelated constructs (e.g., mental health or social function) should be <0.30 .
4	The correlation (or change) with instruments measuring similar constructs should differ by at least 0.10 from the correlation (or change) with instruments measuring related but dissimilar constructs.
5	The correlation (or change) with instruments measuring related constructs should differ by at least 0.10 from the correlation (or change) with instruments measuring unrelated constructs.

3. Results

3.1 Study Selection Results

The search yielded 1054 records. After removing 841 duplicates, 180 were excluded based on inclusion/exclusion criteria. Following full-text review, 22 articles were excluded, leaving 11 studies for inclusion. The flow diagram is presented in Figure 1. Demographic characteristics of the included studies and their populations are shown in Table 3. The 11 articles covered language versions from Germany, India, Sweden, Hong Kong (traditional Chinese), Mainland China (simplified Chinese), Korea, Italy, Sweden (again), Arabic, the Netherlands, and Israel. A summary of measurement property results is provided in Table 4.

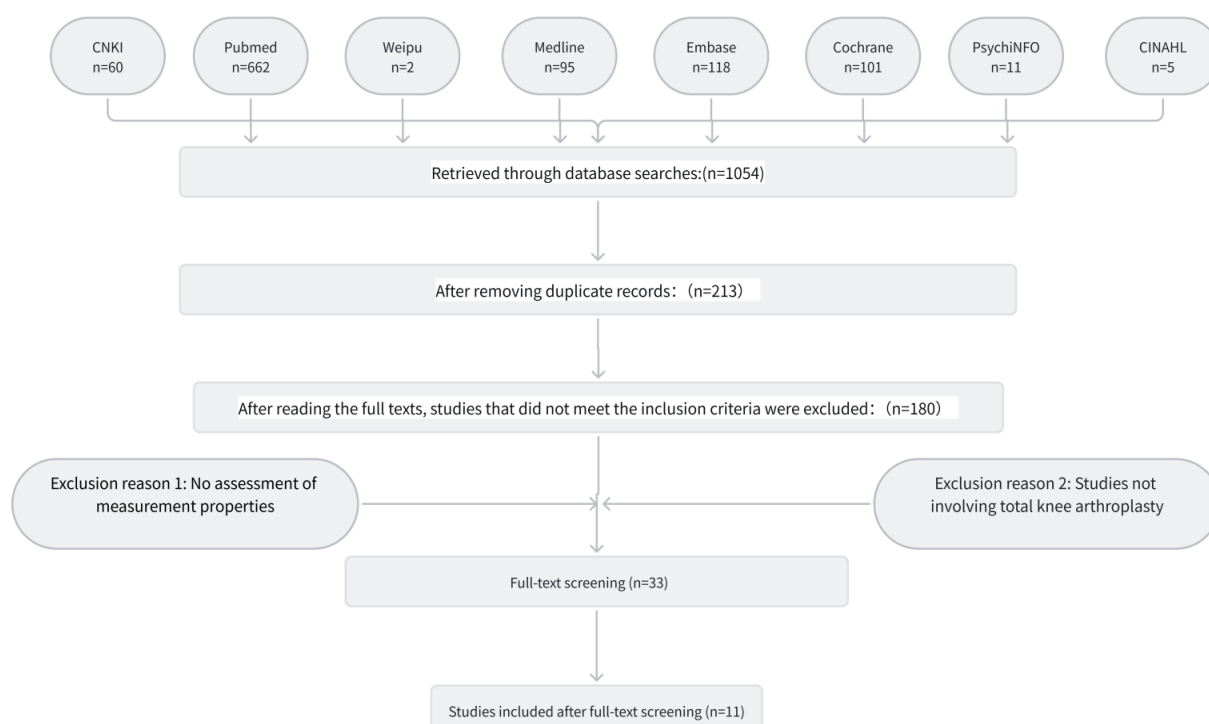


Figure 1: Literature search flow diagram

Table 3: Study and population characteristics

Author	Country; Language	Questionnaire mode	Single or multi-center	Scale	Sample size	Mean age (SD), range	Disease characteristics	Sex distribution (female)	Follow-up
Florian Baumann 2015	Germany; German	Self-administered	Single-center	G-FJS	105	65.2±9.3 (43-85岁)	TKA	57(54.3%)	2 weeks
Tarun Goyal 2020	India; Hindi	Self-administered	Single-center	FJS-12	140	48.68 ± 17.33	TKA	63(44.8%)	2 weeks
Siri HEJBEL 2020	Sweden; Swedish	Self-administered	Single-center	FJS-12	109	69	TKA	55(51%)	2 weeks
Kevin Ki-Wai Ho 2022	Hong Kong; Traditional Chinese	Self-administered	Single-center	FJS	75	NA	TKA	THA:49; TKA:53	2 weeks
Waleed Albishi 2023	Arabic; Arabic	Self-administered	Single-center	FJS	111	61.3±8.2	TKA	90(81.1%)	2 weeks
Shadid MB 2016	Netherlands; Dutch	Self-administered	Single-center	FJS	85	68.5 (9.2; 39-91)	TKA(53%); THA(47%)	TKA(53); THA(49)	2 weeks
Amit Pansky 2021	Israel; Hebrew	Self-administered	Single-center	FJS	110	67.42 ± 7.15	TKA	68(67%)	2 weeks
Shiqi Cao 2017	China; Simplified Chinese	Self-administered	Single-center	SC-FJS	150	68.1 ± 7.4 (47-86岁)	TKA	118(78.7%)	2 weeks
Valerio Sansone 2020	Italy; Italian	Self-administered	Single-center	FJS-12	120	70.0±8.4岁(44-86岁)	TKA	81(67.5%)	2 weeks
Jangyun Lee 2021	Korea; Korean	Self-administered	Single-center	K-FJS	150	71.3 ± 6.5 (57-84)	TKA	124 (82.7)	2 weeks
D. F. Hamilton 2017	UK; English	Self-administered	Single-center	FJS-12	231	69.9±9.6,(36 -91)	TKA	135(58.5%)	2 weeks

*NA: unknown data

Table 4: Results for each measurement property

	Methodological quality	Summary or pooled result	Overall rating of measurement property	No information available
Content validity	Adequate	Relevance adequate (moderate), comprehensiveness inadequate (moderate), comprehensibility inadequate (moderate)	Sufficient	Moderate
Structural validity	Doubtful	No information available	Indeterminate	No information
Internal consistency	Very good	Pooled Cronbach's alpha = 0.932; total N=1378	Sufficient	High
Cross-cultural validity	No information	No information	No information	No information
Reliability	Very good	Pooled ICC = 0.917; total N=1147	Sufficient	High
Measurement error	Doubtful	SDC > MIC	Insufficient	Very low
Hypotheses testing	Inadequate	28% of studies met hypotheses	Insufficient	Moderate
Responsiveness	Doubtful	SRM values insufficient	Insufficient	Very low

3.2 Results of Methodological Quality and Measurement Property Assessment

3.2.1 Content validity

Regarding content validity, the 11 studies were unclear in PROM development about whether experienced moderators or interviewers were employed, whether interviews were based on appropriate topic guides, and whether all group meetings or interviews were recorded and transcribed verbatim. The methodological quality of relevance, comprehensiveness, and comprehensibility of the FJS content was rated as moderate evidence, with the measurement property rated as sufficient.

3.2.2 Structural validity

Among the 11 articles [21–31], only the English version study [31] used confirmatory factor analysis (CFA) and exploratory factor analysis (EFA), with CFI and TLI > 0.99 and RMSEA = 0.075 (> 0.06), indicating good model fit, no methodological flaws, and sufficient sample size; thus, the measurement property was rated as sufficient. The other 10 studies [6–8, 11, 13–18] did not perform CFA or EFA, leading to a methodological quality rating of doubtful and measurement property rating of indeterminate.

3.2.3 Internal consistency

All 11 studies [21–31] reported Cronbach's $\alpha > 0.70$, with a pooled α of 0.932 across 1378 patients. Therefore, methodological quality was rated as very good and the measurement property as sufficient.

3.2.4 Reliability

Eleven studies [21–31] reported reliability; 10 of them [21–30] had ICCs > 0.70, and the pooled ICC across 1147 patients was 0.917. Hence, methodological quality was very good and measurement property sufficient.

3.2.5 Measurement error

In four studies [22, 24, 25, 29], SDC > MIC, while seven studies [21, 23, 26–28, 30, 31] did not provide MIC values. Consequently, methodological quality for measurement error was doubtful, and the measurement property was insufficient.

3.2.6 Cross-cultural validity/Measurement invariance

Cross-cultural validity or measurement invariance of the FJS could not be assessed. Although three studies [22, 28, 32] rated

the measurement property for cross-cultural validity as sufficient, the information remained insufficient.

3.2.7 Criterion validity

None of the 11 studies provided data on a gold standard for the FJS or AUC values, making criterion validity impossible to assess.

3.2.8 Responsiveness

Standardized response means (SRMs) were not reported in any of the 11 studies. Thus, methodological quality was rated as doubtful and measurement property as insufficient.

3.2.9 Floor and ceiling effects

In the Hebrew version [27], floor and ceiling effects were both 0%, indicating low floor/ceiling effects. The German team [25] reported 0% floor effect and no ceiling effect. Although the Korean and English versions [30, 31] did not specifically describe floor effects, ceiling effects were < 15%. Six studies [21, 23, 26–29] reported both floor and ceiling effects < 15%, suggesting low bias. Compared with traditional joint scores such as WOMAC [33], the FJS shows smaller ceiling effects, further supporting its high quality and suitability as a prognostic outcome measure in most countries. Two studies [23, 29] explicitly stated no floor or ceiling effects, reinforcing the cross-cultural applicability of the FJS.

3.2.10 Feasibility

Feasibility is described in Table 5.

Table 5

Feasibility aspect	FJS
Patient comprehensibility	Considered understandable by patients
Clinician comprehensibility	Considered understandable by clinicians
Ease of administration and use	Self-administered, easy to use
Number of items	Brief, 5 items
Completion time	Not specified, assumed ≤ 3 minutes
Required patient ability	Used for > 13 years; patients with knee complaints, generally capable
Ease of standardization	No data available
Ease of scoring	Easy
Copyright	Permission not required
Cost	Free
Equipment needed	Paper or electronic
Usability in different settings	Self-administered or by telephone follow-up
Regulatory approval requirements	Unknown

3.2.11 Hypotheses testing

The structural validity of the FJS was insufficient. Six studies

[6,13–17] tested hypotheses by comparing correlations with other measurement tools (see Table 2). A total of 28 correlations were reported, of which 28% met the pre-defined hypotheses, falling below the 75% standard. Therefore, methodological quality was rated as inadequate, and the measurement property as insufficient.

3.3 Evidence Synthesis and Grading

GRADE synthesis and quality ratings: Content validity of the FJS was rated as sufficient overall, with risk of bias for comprehensiveness, relevance, and comprehensibility rated as doubtful. Moreover, none of the studies described the interview process or assessed the comprehensibility of instructions and response options during development. Consequently, the quality of evidence was downgraded to moderate. Structural validity was indeterminate, precluding a GRADE rating. Internal consistency was sufficient with high-quality evidence. Reliability was sufficient with high-quality evidence. Cross-cultural validity, criterion validity, and measurement error could not be graded due to incomplete information. Hypotheses testing was insufficient, and the evidence was downgraded to moderate due to serious risk of bias. Responsiveness was insufficient, with no AUC reported in any included study; the evidence was downgraded to very low due to very serious risk of bias.

4. Discussion

This study provides the first COSMIN-based systematic review [5] of the FJS, identifying the available evidence on its measurement properties. Importantly, the FJS shows deficiencies in certain measurement properties, including measurement error, structural validity, cross-cultural validity, responsiveness, and criterion validity.

For content validity, moderate-quality evidence indicates that the measurement property is sufficient. However, shortcomings exist in the development process, such as whether experienced moderators or interviewers were employed, whether interviews followed appropriate topic guides, and whether group meetings or interviews were recorded and transcribed verbatim. Standardized interview procedures should be clearly described in the development process to ensure scale accuracy and reliability. Sample selection should include patients across all age groups with adequate sample sizes to enhance representativeness and generalizability.

Across 1378 patients, the pooled Cronbach's α of 0.932 indicates high-quality evidence supporting internal consistency [28], suggesting that the FJS has high applicability for internal consistency after TKA across various countries during localization. The pooled ICC of 0.917 across 1147 patients similarly confirms sufficient reliability.

Regarding responsiveness, Lee et al [18] reported an SRM of 0.67, a low value that does not demonstrate adequate responsiveness. No other studies reported SRM values, precluding further assessment. Given the limited research on FJS responsiveness and lack of sufficient data, future studies should focus on improving responsiveness evaluation, ideally collecting data from patients across all age ranges and

calculating SRM values to more comprehensively assess responsiveness across different patient populations.

Floor or ceiling effects are defined as >15% of participants achieving the highest or lowest possible score. In our systematic review of 11 studies, all reported that the proportion of participants achieving the highest or lowest scores was below 15%. Thus, the scale is suitable for use as a prognostic outcome measure in most countries.

We could not fully evaluate criterion validity or cross-cultural validity/measurement invariance of the FJS. For criterion validity, existing studies did not provide data comparing the FJS with a gold standard or AUC values, key indicators of scale accuracy. Future research should include such comparisons and AUC calculations. Although three studies [7,16,27] described the cross-cultural adaptation process and gave positive ratings, the descriptions remained vague. Furthermore, no differential item functioning (DIF) analysis was performed across subgroups. Therefore, subsequent clinical studies should include multi-group analyses to assess the scale's applicability across different cultural backgrounds. Additionally, the content validity of the FJS has not been firmly established, so caution is needed when interpreting all outcome scores and measurement properties [29] Through this systematic review of measurement properties, we found that the FJS adequately reflects patients' physical function, demonstrating relevance and comprehensiveness of its items. Indeterminate structural validity may interfere with the results of other measurement properties; therefore, its impact should be carefully considered when evaluating and interpreting those properties.

All 10 included studies used the original English version of the FJS and performed cross-cultural adaptation, but the quality of adaptation was generally poor. Some items may not align with the socioeconomic realities of different countries. For example, the simplified Chinese version includes the item "When working in your garden, do you have any sensation in your joint?" [16] – such items may not be fully applicable in countries like China or India, where few people have gardens. Thus, after translation into different languages, higher-quality cross-cultural adaptation is needed. Moreover, some language versions had overly homogeneous samples. For instance, in the traditional Chinese (Hong Kong) version [7], over 80% of participants were elderly, which may not fully represent the FJS score levels of Hong Kong society as a whole and could bias results. Therefore, each national version should be localized according to its own context rather than merely translated; otherwise, some questions may be answered inaccurately or ignored.

Although this study has made progress in scale evaluation, several limitations exist. First, due to limited sample size, we only included articles from various language versions. Resource constraints prevented the inclusion of all potentially relevant literature, which may affect the comprehensiveness of the scale assessment. Second, when applying the COSMIN Risk of Bias checklist and quality criteria to some items, the high subjectivity of these items may have introduced bias. Third, the included studies did not fully evaluate all measurement properties, limiting the generalizability of our conclusions, which should therefore be interpreted cautiously.

Finally, the measurement properties of a scale are dynamic and evolve with cultural, linguistic, temporal, measurement-related, conceptual, target-population, and disease-spectrum changes, requiring continuous accumulation of new evidence.

5. Conclusion

Based on COSMIN standards, this study scientifically supports the use of the FJS in clinical practice while ensuring methodological quality. The results indicate that reporting of measurement properties for this scale is incomplete. Even though high-quality evidence supports sufficient internal consistency and reliability, the assessment of structural validity remains non-standardized. The various cross-culturally adapted versions lack necessary EFA and CFA. Future studies should routinely include EFA and CFA to more precisely define the dimensionality of each version. No floor or ceiling effects were observed. Therefore, the FJS can capture changes in physical function and pain status in TKA patients. Future research should further supplement structural validity, cross-cultural validity, and criterion validity to fully and accurately determine whether the FJS is suitable for use in Chinese TKA patients.

References

- [1] Fischer M, von Eisenhart-Rothe R, Simank H G. Comparable short-term results seen with standard and high-flexion knee arthroplasty designs in european patients[J]. *Journal of Orthopaedics*, 2013, 10(3): 119-122.
- [2] Gagnier J J, Huang H, Mullins M, et al. Measurement properties of patient-reported outcome measures used in patients undergoing total hip arthroplasty: a systematic review[J]. *JBJS Reviews*, 2018, 6(1): e2.
- [3] Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines[J]. *Journal of Clinical Epidemiology*, 1993, 46(12): 1417-1432.
- [4] Calvert M, Blazeby J, Altman D G, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension[J]. *JAMA*, 2013, 309(8): 814-822.
- [5] Prinsen C a. C, Mokkink L B, Bouter L M, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures[J]. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 2018, 27(5): 1147-1157.
- [6] Heijbel S, Naili J E, Hedin A, et al. The forgotten joint score-12 in swedish patients undergoing knee arthroplasty: a validation study with the knee injury and osteoarthritis outcome score (KOOS) as comparator[J]. *Acta Orthopaedica*, 2020, 91(1): 88-93.
- [7] Ho K K W, Chau W W, Lau L C M, et al. Traditional chinese-hong kong version of forgotten joint score-12 (FJS-12) for patients with osteoarthritis of the knee underwent joint replacement surgery: cross-cultural and sub-cultural adaptation, and validation[J]. *BMC musculoskeletal disorders*, 2022, 23(1): 222.
- [8] Behrend H, Giesinger K, Giesinger J M, et al. The "forgotten joint" as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure[J]. *Journal of Arthroplasty*, 2012, 27(3): 430-436.e1.
- [9] Dong Ziyang, Li Yang, Tian Hua. Research progress on the application of Patient Reported Outcomes Assessment Scale in total knee arthroplasty [J]. *Journal of Chinese Academy of Medical Sciences*, 2023, 45(2): 327-333.
- [10] Goyal T, Sethy S S, Paul S, et al. Good validity and reliability of forgotten joint score-12 in total knee arthroplasty in hindi language for indian population[J]. *Knee surgery, sports traumatology, arthroscopy: official journal of the ESSKA*, 2021, 29(4): 1150-1156.
- [11] Adriani M, Malahias M A, Gu A, et al. Determining the validity, reliability, and utility of the forgotten joint score: a systematic review[J]. *Journal of Arthroplasty*, 2020, 35(4): 1137-1144.
- [12] Albishi W, AbuDujain N M, Arafah O, et al. Cross-cultural adaptation, validity and reliability of the arabic version of the forgotten joint score for knee arthroplasty[J]. *Knee Surgery, Sports Traumatology, Arthroscopy: Official Journal of the ESSKA*, 2023, 31(10): 4312-4318.
- [13] Baumann F, Ernstberger T, Loibl M, et al. Validation of the german forgotten joint score (G-FJS) according to the COSMIN checklist: does a reduction in joint awareness indicate clinical improvement after arthroplasty of the knee? [J]. *Archives of Orthopaedic and Trauma Surgery*, 2016, 136(2): 257-264.
- [14] Hamilton D F, Loth F L, Giesinger J M, et al. Validation of the english language forgotten joint score-12 as an outcome measure for total hip and knee arthroplasty in a british population[J]. *The Bone & Joint Journal*, 2017, 99-B (2): 218-224.
- [15] Lohr K N, Aaronson N K, Alonso J, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria[J]. *Clinical Therapeutics*, 1996, 18(5): 979-992.
- [16] Terwee C B, Bot S D M, De Boer M R, et al. Quality criteria were proposed for measurement properties of health status questionnaires[J]. *Journal of Clinical Epidemiology*, 2007, 60(1): 34-42.
- [17] Mokkink L B, De Vet H C W, Prinsen C A C, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures[J]. *Quality of Life Research*, 2018, 27(5): 1171-1179.
- [18] Terwee C B, Prinsen C A C, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study[J]. *Quality of Life Research*, 2018, 27(5): 1159-1170.
- [19] Evaluation of Patient-Reported Outcome Measurement Tools Using a Modified Quantitative Systematic Review Evidence Grading Method.pdf[Z].
- [20] Mokkink L B, Terwee C B, Patrick D L, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study[J]. *Quality of Life Research*, 2010, 19(4): 539-549.
- [21] Heijbel S, Naili J E, Hedin A, et al. The Forgotten Joint Score-12 in Swedish patients undergoing knee arthroplasty: a validation study with the Knee Injury and Osteoarthritis Outcome Score (KOOS) as comparator[J]. *Acta Orthopaedica*, 2020, 91(1): 88-93.

- [22] Ho K K W, Chau W W, Lau L C M, et al. Traditional Chinese-Hong Kong version of Forgotten Joint Score-12 (FJS-12) for patients with osteoarthritis of the knee underwent joint replacement surgery: cross-cultural and sub-cultural adaptation, and validation[J]. *BMC musculoskeletal disorders*, 2022, 23(1): 222.
- [23] Albishi W, AbuDujain N M, Arafah O, et al. Cross-cultural adaptation, validity and reliability of the Arabic version of the Forgotten Joint Score for knee arthroplasty[J]. *Knee surgery, sports traumatology, arthroscopy: official journal of the ESSKA*, 2023, 31(10): 4312-4318.
- [24] Goyal T, Sethy S S, Paul S, et al. Good validity and reliability of forgotten joint score-12 in total knee arthroplasty in Hindi language for Indian population. [J]. *Knee surgery, sports traumatology, arthroscopy: official journal of the ESSKA*, 2021, 29(4): 1150-1156.
- [25] Baumann F, Ernstberger T, Loibl M, et al. Validation of the German Forgotten Joint Score (G-FJS) according to the COSMIN checklist: does a reduction in joint awareness indicate clinical improvement after arthroplasty of the knee? [J]. *Archives of Orthopaedic and Trauma Surgery*, 2016, 136(2): 257-264.
- [26] Shadid M B, Vinken N S, Marting L N, et al. The Dutch version of the Forgotten Joint Score: test-retesting reliability and validation[J]. *Acta Orthopaedica Belgica*, 2016, 82(1): 112-118.
- [27] Pansky A, Bar-Ziv Y, Tamir E, et al. Reliability and validity of the Hebrew version of the forgotten joint score for assessing the outcomes of total knee arthroplasty[J]. *Arthroplasty (London, England)*, 2021, 3(1): 27.
- [28] Cao S, Liu N, Han W, et al. Simplified Chinese version of the Forgotten Joint Score (FJS) for patients who underwent joint arthroplasty: cross-cultural adaptation and validation[J]. *Journal of Orthopaedic Surgery and Research*, 2017, 12(1): 6.
- [29] Sansone V, Fennema P, Applefield R C, et al. Translation, cross-cultural adaptation, and validation of the Italian language Forgotten Joint Score-12 (FJS-12) as an outcome measure for total knee arthroplasty in an Italian population[J]. *BMC Musculoskeletal Disorders*, 2020, 21(1): 23.
- [30] Lee J, Lim S H, Ro D H, et al. Translation and Validation of the Korean Version of the Forgotten Joint Score[J]. *Clinics in Orthopedic Surgery*, 2021, 13(4): 482-490.
- [31] Hamilton D F, Loth F L, Giesinger J M, et al. Validation of the English language Forgotten Joint Score-12 as an outcome measure for total hip and knee arthroplasty in a British population[J]. *The Bone & Joint Journal*, 2017, 99-B(2): 218-224.
- [32] Thomsen M G, Latifi R, Kallemose T, et al. Good validity and reliability of the forgotten joint score in evaluating the outcome of total knee arthroplasty[J]. *Acta Orthopaedica*, 2016, 87(3): 280-285.
- [33] Thomsen M G, Latifi R, Kallemose T, et al. Good validity and reliability of the forgotten joint score in evaluating the outcome of total knee arthroplasty: A retrospective cross-sectional survey-based study[J]. *Acta Orthopaedica*, 2016, 87(3): 280-285.