Assessment of cross-cultural adaptations and patient-reported outcome measures relevant to shoulder disorders in Turkish: A systematic review using the COSMIN methodology

Cagdas Isiklar[1,2]*, Gamze Cagla Sirma[3], Elif Turgut[1]

1 Department of Sports Physiotherapy, Faculty of Physical Therapy and Rehabilitation, Hacettepe University, Ankara, Türkiye, 2 Department of Physiotherapy and Rehabilitation, Faculty of Health Sciences, Fenerbahce University, Istanbul, Türkiye, 3 Department of Occupational Therapy, Faculty of Health Sciences, Fenerbahce University, Istanbul, Türkiye

* cagdasisiklar@gmail.com, cagdas.isiklar@fbu.edu.tr

Abstract

Background

There are many shoulder assessment outcome measures in the literature that have been studied for validity and reliability. However, there is no study examining the Turkish-adapted patient outcome measures (PROMs) on the shoulder according to the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist. In addition, there is a small number of studies that carry out this examination on the shoulder internationally. Determining the most appropriate questionnaires for clinical use will also be an important guide in patient evaluation by filling the gap in both literature and clinical aspects.

Objective

Our aim in this study is to identify the valid and reliable Turkish scales used to evaluate shoulder pain and disability, to reveal how compatible these scales are with the aspects of study quality and psychometric quality according to COSMIN criteria list.

Methods

A systematic search was performed in the following electronic databases: MED-LINE, Web of Science (WOS), EMBASE, SCOPUS, and ULAKBIM were searched from inception to September 2024.

Results

A total of 2890 articles were found in the databases. Four articles were included additionally without a search. Then, after eliminating the duplicates, 2231 articles were included based on the title and abstract search. 63 articles were included in full-text

review. After the full-text review, according to our criteria, 29 articles were included in the systematic review. According to the assessment by the COSMIN checklist, 28 articles received the final classification of "inadequate". Only 1 article (UCLA-Tr) received the final classification of "adequate".

Conclusion

There is a wide range of Turkish PROMs available for the shoulder, with only one fully meeting COSMIN methodology criteria. However, this does not imply that other PROMs lack clinical utility. Future studies should focus on refining PROMs by incorporating MCID and PASS scores and prioritizing their responsiveness to enhance their clinical relevance. Additionally, variations in PROM performance across different patient populations highlight the need for context-specific evaluations to ensure their applicability in diverse clinical settings.

Trial registration

PROSPERO registration number: CRD42022298425.

Introduction

The prevalence of shoulder pain in the general population has been reported to range between 0.7% and 55.2%, with a median value of 16.0% [1]. Rotator cuff-related shoulder pain, tendinopathies, instabilities, and frozen shoulder are the common shoulder disorders and result in a wide range of functional disability, performance impairments, and loss of quality of life [2–5]. Recently, the advances in the management of shoulder disorders have been made through more comprehensive and accurate assessment tools in rehabilitation [6].

In addition to clinical evaluation and imaging methods, measuring treatment outcomes is an important aspect of the evaluation of patient perspective about their health status. For this purpose, patient-reported outcomes measures (PROMs) are commonly used to provide an objective measure of patients' views of their subjective symptoms, such as quality of life, functional status, or satisfaction with treatment and disability.

More than 40 instruments are available for both generic and disease-specific use in shoulder disorders [7–10]. When choosing the best PROM developed for a specific purpose, such as shoulder disability or pain, for use in clinical practice or research, it is necessary to evaluate and present the psychometric properties of existing questionnaires such as validity, reliability, and responsiveness. Therefore, the purpose and importance of measurement features of outcome measurement tools according to the COSMIN taxonomy reflect the relevance and evaluation for any measurement tool used in any application. The predictive validity of a psychometric questionnaire is therefore the essential factor in predicting a person's attitude [11].

The Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) group have developed a consensus-based standard for assessing the quality of studies on measurement properties and provided a methodology for conducting systematic reviews of PROMs [12–14]. To determine the methodological quality based on COSMIN, the characteristics of the PROMs are evaluated against the COSMIN checklist steps.

Previous reviews of the psychometric properties of shoulder PROMs suggested that the Western Ontario Rotator Cuff (WORC) Index, the Disabilities of the Arm, Shoulder and Hand Questionnaire (DASH), the Shoulder Pain and Disability Index (SPADI), and the Simple Shoulder Test (SST) had good evidence according to GRADE approach [15,16]. According to the GRADE approach, it is essential for these surveys to provide robust evidence. Evaluating the GRADE and COSMIN approaches in conjunction offers a comprehensive understanding of the highest level of evidence. While the GRADE approach assesses the quality of evidence for clinical decisions and recommendations, focusing on aspects such as study design, risk of bias, consistency, precision, and directness [17], the COSMIN approach emphasizes the evaluation of measurement properties of health instruments, including reliability, validity, responsiveness, and methodological rigor [18]. Together, these frameworks complement each other, providing a more holistic evaluation of evidence. About measuring internal consistency this is the extent to which items within a scale measure the same underlying construct. It indicates the homogeneity of the items. Typically assessed using Cronbach's alpha. A value between 0.7 and 0.9 is generally considered acceptable, though this depends on the context [19]. For reliability it is extent to which an instrument consistently measures a construct across time, raters, or conditions. Test-Retest Reliability: Consistency of results when the same instrument is applied at different times. Inter-Rater Reliability: Agreement between different observers using the same tool. Intra-Rater Reliability: Consistency of the same observer measuring repeatedly. Metrics like intraclass correlation coefficient (ICC) or kappa statistics can be used for measure [20]. Lastly, validity extent to which an instrument measures what it is intended to measure. Content Validity: Whether the instrument covers all relevant aspects of the construct. Construct Validity: The degree to which the instrument aligns with theoretical expectations (e.g., correlations with related constructs). Criterion Validity: The extent to which the instrument correlates with a gold standard or an external criterion. Correlations, factor analysis, or comparison with established measures can be used for assess [21]. Almost all of the currently available instruments have been developed in English-speaking countries. To be able to administer a questionnaire in a culture or language other than the culture it was developed for or the original language, an appropriate cross-cultural adaptation process must be performed so that it retains the same meaning as the original [22]. Thus, cross-culturally adapted instruments can be used for assessment and reporting only after possible misinterpretations that may occur due to cultural or lifestyle differences are eliminated.

There are several PROMs related to shoulder diseases that have been translated into Turkish [23–51]. For Turkish speakers, the Western Ontario Shoulder Instability Index – Turkish version (WOS-Tr) is recommended for glenohumeral instability, the Rotator Cuff-Quality of Life Scale – Turkish version (RCQOL-Tr) for rotator cuff disease, arm, shoulder, and hand disabilities [52] and the Shoulder Pain and Disability Index – Turkish version (SPADI-Tr) for non-specific shoulder pain [16]. However, there is a need for systematic reviews reporting a standardized evaluation of the psychometric properties of Turkish PROMs for shoulder diseases. Therefore, the present study aimed to systematically review the study quality and psychometric quality of Turkish PROMs for shoulder diseases using the COSMIN methodology to provide data for Turkish-speaking researchers and clinicians about which PROM is best for a specific purpose.

Methods

Study selection

Systematic searches were performed in the following electronic databases to identify available Turkish language tools for assessing shoulder pain and disability and to demonstrate their compatibility with the COSMIN criteria list: MED-LINE, Web of Science, EMBASE, SCOPUS, and ULAKBIM. The literature search was tailored to each database and based on the protocol suggested by the COSMIN group. The search terms and the Boolean operators used in the databases

included "shoulder joint", "shoulder related disorders", "instruments", and "Turkish versions". Search was performed both in English and Turkish. There was no limitation in terms publication date or language limitation. The last search was performed in September 2024. All articles were scanned on Google Scholar using the search terms mentioned above. Details of the search strategy and related criteria are provided in S6 and S7 Tables.

Inclusion criteria of our study: Studies examining the psychometric properties of instruments originally developed in another language and translated and adapted into Turkish for the assessment of shoulder joint complex or upper extremity dysfunctions (Guidelines and rating criteria for the process of cross-cultural adaptation of self-report measures are detailed in S8 Table), all PROMs including self-report tools or interview-based questionnaires that use objective measures, only full-text articles. Exclusion criteria of our study: Instruments developed for specific groups whose primary complaint did not concern shoulder-related musculoskeletal disorders (e.g., wheelchair users and patients with cancer), And also study design is based on prognostic studies, systematic reviews, meta-analyses, abstracts from conferences, books, and theses/dissertations.

Assessment of the methodological quality of eligible studies and data extraction

Included articles have been evaluated in terms of translation and adaptation across cultures. Results are presented according to the standards and translation guidelines developed by Guillemin et al. [53] and/or Beaton et al. [54], which involves 5 steps for cross-cultural translation and adaptation as follows: (1) first translation; (2) synthesis; (3) back translation; (4) review by a committee of experts in the field; and, (5) pretesting the final version.

Data regarding the translation and cross-cultural adaptation were extracted in order to assess the design of these studies. In addition, data on the measurement properties of the COSMIN checklist were extracted for each study [55]. After that, the cross-cultural adaptation and translation methods of each study were classified according to the COSMIN methodology. Two reviewers (C.I. & G.C.S.) independently performed the assessment of the methodological quality for each criterion of the COSMIN checklist. The COSMIN checklist was used to evaluate general methodological quality. Each criterion was rated independently by the two reviewers to ensure an unbiased assessment process. Discrepancies between reviewers were addressed through a structured discussion process. In cases where consensus could not be reached, a third reviewer (E.T.) resolved the conflicts. The final ratings were derived through a consensus process, considering input from all three reviewers. To facilitate the review process and systematically manage conflicts, the Rayyan platform (https://www.rayyan.ai/) was utilized. This platform allowed for blinded independent assessments by reviewers and provided a structured environment for resolving disagreements efficiently.

The checklist contains nine boxes with standards for good methodological quality of studies on nine different measurement properties. The quality of each step is marked positive (+) when the process is performed in accordance with quality criteria; it is marked doubtful (?) when the definition of the method is unclear and there are insufficient quantity of translators and/or back translators; or zero (0) when there is not enough information to evaluate each step [13,56].

The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach (S8 Table) is a systematic and transparent approach to grading the precision of evidence in systematic reviews and clinical practice guidelines and to develop and determine the strength of clinical practice recommendations. The quality rating of each measurement feature can be added directly to the relevant table. If the results of some studies are ignored when condensing the evidence and determining the overall rating of the pooled or summarized outcome for a measurement feature, those studies should also be ignored in determining the quality of the evidence [57]. In this study, we identified "high-quality studies" using the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach. According to GRADE, specifically, these studies were identified based on their ability to deliver high-quality evidence regarding validity, consistency, and reliability, as assessed through the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach.

In summary, we have used three sets of evaluation criteria—COSMIN, GRADE, and measurement translation suggestions—to assess the measures comprehensively: The COSMIN approach was used to evaluate the methodological quality

of studies related to the development and validation of the instruments. The GRADE approach was applied to assess the overall quality of evidence for each instrument, focusing on validity, reliability, consistency, and relevance to our research question. The measurement translation suggestions were included to address the practical aspects of adapting measures for cross-cultural use, ensuring conceptual and linguistic equivalence.

Results

A total of 2890 records were identified in the search. In addition, four articles were added with hand search. Based on the titles and abstracts, 63 articles were selected. Thirty-four articles that met the selection criteria were excluded after assessment due to missing duplicates and study design. Finally, 29 articles were included in this study (Fig 1). During the full-text screening, the reviewers had 93.1% agreement. The conflict on the remaining 6.9% (2 articles) was resolved after a meeting.

The questionnaires translated and adapted to Turkish were: the Western Ontario Osteoarthritis of the Shoulder Index (WOOS) [23]; the University of California Los Angeles (UCLA) Shoulder Scale [24]; the Shoulder Rating Questionnaire (SRQ) [25]; L'Insalata Shoulder Rating Questionnaire (LSRQ) [26]; the Specific Acromioclavicular Score (SACS) [27]; The Long Head of Biceps Score (LHB) [28]; the Oxford Shoulder Instability Score (OSIS) [29]; the Kerlan-Jobe Orthopaedic Clinic Shoulder and Elbow Score (KJOC-SES) [30]; the Penn Shoulder Scale (PSS) [31]; the Western Ontario Shoulder Instability Index (WOSI) [32]; the Modified Constant Murley Score (CMS) [33]; the Rotator Cuff Quality of Life (RC-QOL) [45]; the Rotator Cuff-Quality of Life Scale (RC-QoLS) [46]; the Upper LIMB Functional Index (ULFI) [49]; the Upper Extremity Functional Index (UEFI) [35]; the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASES) [36]; the Milliken Activities of Daily Living Scale (MAS) [37]; the Quick Disability of the Arm, Shoulder, and Hand (Q-DASH) [38]; the Oxford Shoulder Score (OSS) [58]; the Shoulder Pain and Disability Index (SPADI) (adaptation, reliability and validity study) [34]; the Shoulder Pain and Disability Index (SPADI) (validation study for Turkish women) [40]; the Simple Shoulder Test (SST) [41]; the Shoulder Disability Questionnaire (SDQ) [43]; the Disability of the Arm, Shoulder, and Hand (DASH) (validity study for workers) [42]; the Disability of the Arm, Shoulder, and Hand (DASH) (adaptation, reliability and validity study) [44]; Munich Shoulder Questionnaire (validity and reliability study) [50]; the Western Ontario Rotator Cuff Index (WORC) [48]; Nottingham Clavicle Score (validity and verification study) [51]; and a comparison of the responsiveness of SDQ, SPADI, and WORC index (Responsiveness study) [47] (S1 Table).

The 5 steps of the translation process were performed in 23 studies included [23,24,26–28,30–39,43–46,48–51] (S2 Table). The results of high-quality studies are considered when determining the overall rating, the rating of the evidence is determined only by the high-quality studies (in which case we do not downgrade for the risk of bias) In the first assessment, the assessors had 88.89% agreement (32 out of 36 items). After further discussion, they reached 100% consensus. For the quality of evidence analysis, there was 100% agreement between the two reviewers (S3 Table). S4 Table presents the detailed analysis of the risk of bias and quality appraisal based on COSMIN risk of bias checklist, updated criteria for good measurement properties about questionnaires/scales.

Twenty-six questionnaires assessed internal consistency [23–26,28–43,45,46,48–51]; twenty-seven assessed criterion validity [23–45,48–51], and twenty-four analyzed reproducibility [23–28,30–33,35–46,48,49]. There is a lack of analysis among the questionnaires, mainly in relation to the properties floor/ceiling effect and responsiveness, which were verified only in nine [24,26,27,30,32,33,35,36,39] and two articles [24,50], respectively.

The SF-36 questionnaire (mental and quality of life), SPADI, and DASH was the most used comparative for validity assessment among the studies. In addition to these questionnaires, WORC, ASES, UEFI, OSS, PSS, Constant Score, and SF-12 were seen as the questionnaires used for comparison. Two of the 29 articles did not apply questionnaires/scales for comparison [46,47].

According to the assessment by the COSMIN scale, 28 articles received the final classification of "inadequate". Only 1 article [24] received the final classification of "adequate" (**S5 Table**). Also, in the inclusion of each study and review of the studies according to the COSMIN checklist, three authors independently completed this process.
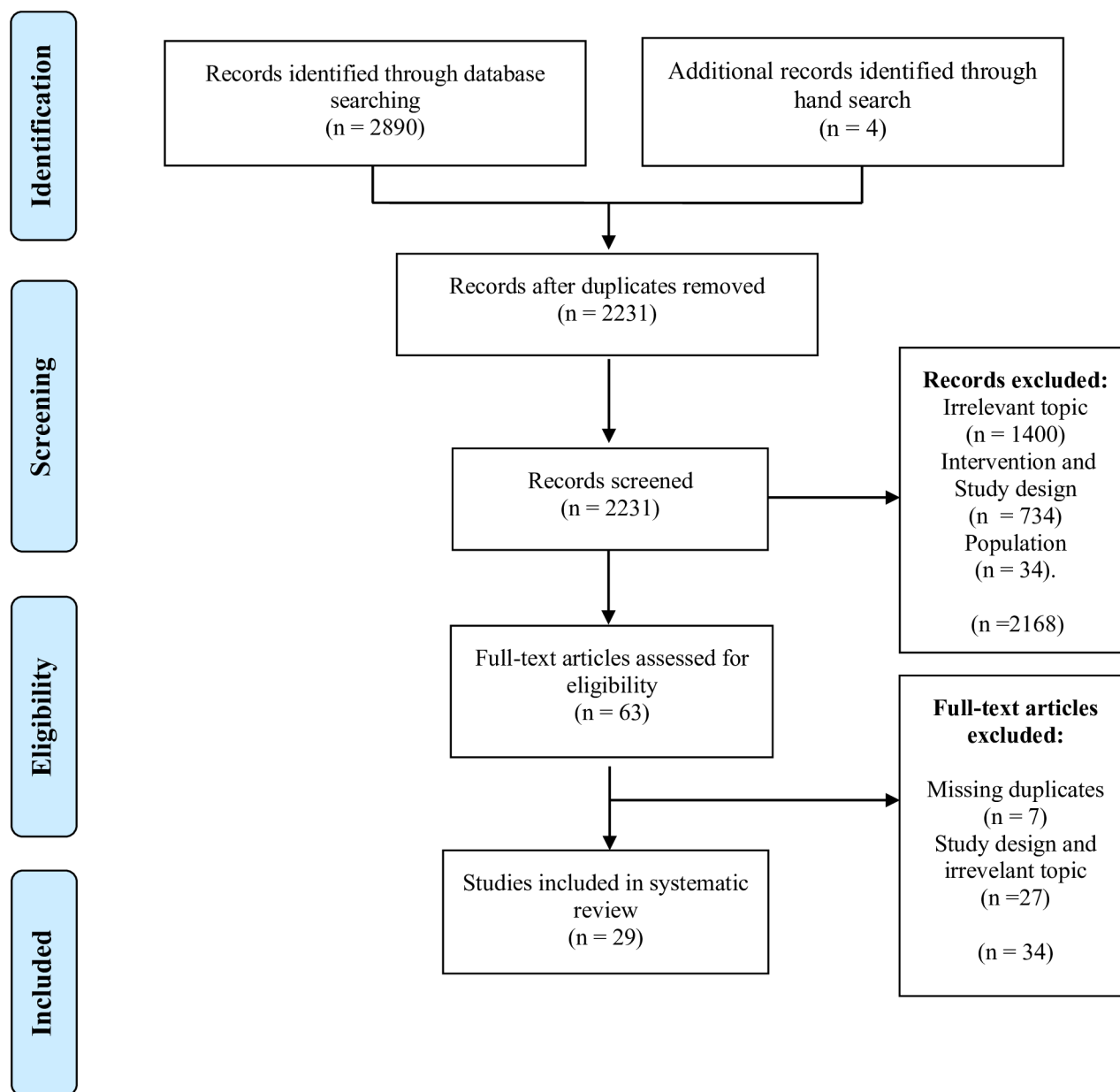
Fig 1. PRISMA flowchart. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-analyses.

Discussion

This systemic review aimed to evaluate the available literature regarding the psychometric properties of Turkish PROMs for shoulder diseases using the COSMIN methodology. A total of 28 distinct PROMs, out of 29 examining shoulder disorders were adapted to Turkish.

Regarding the methodological quality of the included studies according to the COSMIN checklist, there is mostly moderate or poor evidence available. The findings of this study showed that, OSS-Tr, ASES-Tr, CMS-Tr, were found proficient PROMs according to the COSMIN methodology and the UCLA-Tr was found the most appropriate PROM in this study.

However, commonly used questionnaires in Turkish provide low-level evidence (S5 Table) regarding rotator cuff disease, glenohumeral instability, subacromial impingement syndrome, acromioclavicular joint instability, or other shoulder-specific upper extremity according to the COSMIN methodology.

This is the first review in the literature evaluating the available cross-cultural adaptations of shoulder PROMs using COSMIN methodology. Previously, shoulder-related PROMs translated into Portuguese were reviewed without using the COSMIN checklist [59]. It was stated that DASH and WORC are the most appropriately developed and tested questionnaires in the study conducted in Portugal [59]. In the present study, the UCLA-Tr questionnaire was determined as the most appropriate questionnaire. Most of the adapted PROMs have not been investigated for responsiveness. In addition, the clinical use of these PROMs and their suitability for patients have not been demonstrated by the minimal clinically important difference (MCID) or the patient acceptable symptom state (PASS) scores [60]. Future adapted PROMs should take these shortcomings into account.

The COSMIN checklist has been previously applied for the assessment of the original shoulder PROMs. Longo et al. investigated the methodological quality of studies in terms of the psychometric properties of PROMs for rotator cuff disease and recommended the use of WORC and RC-QOL [61]. In their review, Huang et al. deduced that the WORC had the highest ratings, followed by the DASH, SPADI, and SST [15]. Both of the two reviews have suggested that there is a need for higher quality methodological studies to evaluate the characteristics of all instruments identified in rotator cuff and shoulder pathologies. In a systematic review conducted by Villegas et al., which examined outcome measures utilized for assessing shoulder functionality, it was highlighted that a wide variety of functional assessment tools have been employed across different shoulder injuries. While this diversity enhances the range of options available for clinical practice, the authors underscored the critical need for developing high-quality outcome measures that comprehensively address diverse methodological and clinical requirements [62]. In this study, it was revealed that WORC-Tr or RC-QOL-Tr scales are available instruments for rotator cuff diseases, although they were not found to be adequate in quality according to the COSMIN methodology. It will be clinically beneficial to examine the clinical significance in both Turkish and original versions, to follow similar methods in the original patient outcome measures, and to reveal these in systematic reviews.

The main problem is that most of the studies evaluated in this review have not used any checklists for methodological quality. When designing a study evaluating the psychometric properties of a PROM, applying a checklist with the necessary standards and statistical methods will improve the methodological quality of the study [18]. In the literature, there are several methodological guidelines to evaluate the quality of PROMs, such as the Evaluating the Measurement of Patient-Reported Outcomes tool [63], the COSMIN checklist [18], and the checklist developed by Francis et al. [64]. In future studies evaluating the psychometric properties of PROMs, it is recommended to use a guide of some advantages and disadvantages suggested by Mokkink et al. [12]. Higher quality studies are needed to assess all relevant psychometric properties of current PROMs in Turkish, particularly content validity and construct validity, in order to gather stronger evidence. In addition, more research is needed to evaluate the psychometric properties of PROMs with other common musculoskeletal shoulder disorders (e.g., adhesive capsulitis).

While our initial conclusions focused on measures meeting 100% of the COSMIN criteria, we recognize that this approach may overlook the potential value of tools that, while not fully meeting all criteria, demonstrate promising properties in other aspects. For instance, the modified CMD-tr, despite lacking evidence on responsiveness, showed strong performance across most evaluated criteria. This suggests its potential utility in clinical and research settings. We also acknowledge that the absence of a detailed discussion on floor and ceiling effects does not necessarily render a measure inadequate. These effects can vary significantly depending on the population studied and may be less critical than other aspects such as validity and reliability. In this context, other PROMs used in Turkish clinical settings should not be dismissed as inadequate solely based on our findings. Instead, their classification as inadequate reflects their performance against the specific COSMIN-based criteria applied in this study. To better guide future research and practice, we propose evaluating PROMs along a continuum, highlighting their strengths, identifying areas requiring further investigation, and

considering their relevance to specific populations and contexts. This approach could provide a more nuanced and practical framework for selecting and improving measurement tools.

This systematic review had three main limitations. First, theses with unpublished data on the measurement properties of the described instruments could also be included. Second, the findings of this review were not categorized by any specific population or disease and only covered the adult population. Finally, this review lacks the gray literature containing footnotes produced outside of traditional articles and distribution channels.

Conclusion

There is a wide variety of Turkish PROMs for the shoulder, and only one specific scale was demonstrated to fit all of the parameters defined according to the COSMIN methodology. The findings of the current study regarding the psychometric properties of PROMs and the quality of the studies should be assessed in future studies using MCID and PASS scores. Furthermore, special attention should be given to improving the responsiveness of PROMs, as this is a critical aspect of their utility in clinical practice. It should also be noted that the performance of these questionnaires may vary across different patient populations, emphasizing the importance of evaluating their applicability and relevance in diverse clinical contexts. Additionally, it is important to highlight that the classification of PROMs as inadequate based on COSMIN criteria does not necessarily imply that they are clinically unusable. The definitions of inadequacy in this study are rooted in methodological parameters, which may differ from the practical utility of these tools in real-world clinical settings.

Supporting information

S1 Table. Characteristics of the included PROMs evaluated in the systematic review.
(DOCX)

S2 Table. Cross-cultural adaptation steps based on guideline recommendations.
(DOCX)

S3 Table. GRADE assessment of the quality of evidence for measurement properties.
(DOCX)

S4 Table. COSMIN Risk of Bias ratings for each included PROM.
(DOCX)

S5 Table. Summary of psychometric properties and COSMIN quality assessment.
(DOCX)

S6 Table. English and Turkish search strategies used in the review process.
(DOCX)

S7 Table. Criteria and rating system for cross-cultural adaptation of self-report measures.
(DOCX)

S8 Table. Overview of the GRADE framework and downgrading rules.
(DOCX)

S1 Checklist. PRISMA 2020 checklist (2 and 3).
(DOCX)

Acknowledgments

Author contributions

Conceptualization: Cagdas Isiklar, Gamze Cagla Sirma.

Supervision: Elif Turgut.

Writing – original draft: Cagdas Isiklar, Gamze Cagla Sirma.

Writing – review & editing: Elif Turgut.

References

1. Lucas J, van Doorn P, Hegedus E, Lewis J, van der Windt D. A systematic review of the global prevalence and incidence of shoulder pain. BMC Musculoskelet Disord. 2022;23(1):1073. https://doi.org/10.1186/s12891-022-05973-8 PMID: 36476476

2. Bennell K, Wee E, Coburn S, Green S, Harris A, Staples M, et al. Efficacy of standardised manual therapy and home exercise programme for chronic rotator cuff disease: randomised placebo controlled trial. BMJ. 2010;340:c2756. https://doi.org/10.1136/bmj.c2756 PMID: 20530557

3. Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:f167. https://doi.org/10.1136/bmj.f167 PMID: 23358487

4. Toprak M, Erden M. Sleep quality, pain, anxiety, depression and quality of life in patients with frozen shoulder1. J Back Musculoskelet Rehabil. 2019;32(2):287–91. https://doi.org/10.3233/BMR-171010 PMID: 30347591

5. Akyol Y. Shoulder muscle strength in patients with subacromial impingement syndrome: its relationship with duration of quality of life and emotional status. Turk J Phys Med Rehab. 2013;16(3):176–81. https://doi.org/10.4274/tftr.59837

6. Doiron-Cadrin P, Lafrance S, Saulnier M, Cournoyer É, Roy J-S, Dyer J-O, et al. Shoulder Rotator Cuff Disorders: A Systematic Review of Clinical Practice Guidelines and Semantic Analyses of Recommendations. Arch Phys Med Rehabil. 2020;101(7):1233–42. https://doi.org/10.1016/j.apmr.2019.12.017 PMID: 32007452

7. Harvie P, Pollard TCB, Chennagiri RJ, Carr AJ. The use of outcome scores in surgery of the shoulder. J Bone Joint Surg Br. 2005;87(2):151–4. https://doi.org/10.1302/0301-620x.87b2.15305 PMID: 15736732

8. Haywood KL. Patient-reported outcome I: measuring what matters in musculoskeletal care. Musculoskeletal Care. 2006;4(4):187–203. https://doi.org/10.1002/msc.94 PMID: 17117444

9. Roe Y, Soberg HL, Bautz-Holter E, Ostensjo S. A systematic review of measures of shoulder pain and functioning using the International classification of functioning, disability and health (ICF). BMC Musculoskelet Disord. 2013;14:73. https://doi.org/10.1186/1471-2474-14-73 PMID: 23445557

10. Wylie JD, Beckmann JT, Granger E, Tashjian RZ. Functional outcomes assessment in shoulder surgery. World J Orthop. 2014;5(5):623–33. https://doi.org/10.5312/wjo.v5.i5.623 PMID: 25405091

11. Schimidt F, Hunter J. The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. Psychol Bull. 1998;124(2):262.

12. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539–49. https://doi.org/10.1007/s11136-010-9606-8 PMID: 20169472

13. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147–57. https://doi.org/10.1007/s11136-018-1798-3 PMID: 29435801

14. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21(4):651–7. https://doi.org/10.1007/s11136-011-9960-1 PMID: 21732199

15. Huang H, Grant JA, Miller BS, Mirza FM, Gagnier JJ. A Systematic Review of the Psychometric Properties of Patient-Reported Outcome Instruments for Use in Patients With Rotator Cuff Disease. Am J Sports Med. 2015;43(10):2572–82. https://doi.org/10.1177/0363546514565096 PMID: 25622986

16. Thoomes-de Graaf M, Scholten-Peeters GGM, Schellingerhout JM, Bourne AM, Buchbinder R, Koehorst M, et al. Evaluation of measurement properties of self-administered PROMs aimed at patients with non-specific shoulder pain and "activity limitations": a systematic review. Qual Life Res. 2016;25(9):2141–60. https://doi.org/10.1007/s11136-016-1277-7 PMID: 27039305

17. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924–6. https://doi.org/10.1136/bmj.39489.470347.AD PMID: 18436948

18. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10:22. https://doi.org/10.1186/1471-2288-10-22 PMID: 20298572

19. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ. 2011;2:53–5. https://doi.org/10.5116/ijme.4dfb.8dfd PMID: 28029643

20. Streiner D, Norman G, Cairney J. Health measurement scales: A practical guide to their development and use. 5th ed. Oxford University Press; 2024.

21. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar A-R. Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. J Caring Sci. 2015;4(2):165–78. https://doi.org/10.15171/jcs.2015.017 PMID: 26161370

22. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. J Clin Epidemiol. 2015;68(4):435–41. https://doi.org/10.1016/j.jclinepi.2014.11.021 PMID: 25698408

23. Özal Ş, Atalay Güzel N, Kaptan AY, Eren TK, Kafa N. Validity and Reliability of the Turkish Version of Western Ontario Osteoarthritis of the Shoulder Index. Jt Dis Relat Surg. 2021;32(2):497–503. https://doi.org/10.52312/jdrs.2021.141 PMID: 34145829

24. Büyükdoğan K, Koyuncu Ö, Aslan L, Çelik D, Demirhan M. Translation, cross-cultural adaptation, reliability, and validity of Turkish version of the university of California Los Angeles (UCLA) shoulder scale into Turkish. Disabil Rehabil. 2022;44(17):4871–8. https://doi.org/10.1080/09638288.2021.1914754 PMID: 33909532

25. Çiftçi B, Uzunkulaoğlu A, Öke Topçu D. Validity and reliability of the Turkish version of the shoulder rating questionnaire in patients with shoulder pain. Acta Orthop Traumatol Turc. 2021;55(3):208–12. https://doi.org/10.5152/j.aott.2021.191068 PMID: 34100360

26. Yaşar P, Çıtak Karakaya İ, Karakaya MG. Psychometric properties of the Turkish version of L'Insalata Shoulder Rating Questionnaire. Disabil Rehabil. 2022;44(16):4493–9. https://doi.org/10.1080/09638288.2021.1906958 PMID: 33789059

27. Oğuzkaya S, Mısır A, Kızkapan TB, Uzun E, Özçamdallı M, Önder M. Translation, cultural adaptation, and validation of the Turkish-specific acromioclavicular score. Acta Orthop Traumatol Turc. 2021;55(5):406–9. https://doi.org/10.5152/j.aott.2021.20287 PMID: 34730526

28. Najafov E, Özal Ş, Kaptan AY, Ulucaköy C, Kanatlı U, Ataoğlu B, et al. Validity and Reliability of the Turkish Version of LHB Score. J Sport Rehabil. 2020;30(1):30–6. https://doi.org/10.1123/jsr.2019-0364 PMID: 32131050

29. Sonmezer E, Yosmaoglu HB, Doğan CD. The reliability and validity of the Turkish version of the oxford shoulder instability score. Disabil Rehabil. 2020;42(2):261–6. https://doi.org/10.1080/09638288.2018.1497717 PMID: 30326748

30. Turgut E, Tunay VB. Cross-cultural adaptation of Kerlan-Jobe Orthopaedic Clinic shoulder and elbow score: Reliability and validity in Turkish-speaking overhead athletes. Acta Orthop Traumatol Turc. 2018;52(3):206–10. https://doi.org/10.1016/j.aott.2018.02.007 PMID: 29530478

31. Hazar Kanik Z, Gunaydin G, Pala OO, Sozlu U, Alkan ZB, Citaker S, et al. Translation, cultural adaptation, reliability, and validity of the Turkish version of the Penn Shoulder Score. Disabil Rehabil. 2018;40(10):1214–9. https://doi.org/10.1080/09638288.2017.1284905 PMID: 28637131

32. Basar S, Gunaydin G, Hazar Kanik Z, Sozlu U, Alkan ZB, Pala OO, et al. Western Ontario Shoulder Instability Index: cross-cultural adaptation and validation of the Turkish version. Rheumatol Int. 2017;37(9):1559–65. https://doi.org/10.1007/s00296-017-3716-3 PMID: 28429046

33. Çelik D. Turkish version of the modified Constant-Murley score and standardized test protocol: reliability and validity. Acta Orthop Traumatol Turc. 2016;50(1):69–75. https://doi.org/10.3944/AOTT.2016.14.0354 PMID: 26854052

34. Bumin G, Tüzün EH, Tonga E. The Shoulder Pain and Disability Index (SPADI): Cross-cultural adaptation, reliability, and validity of the Turkish version. BMR. 2008;21(1):57–62. https://doi.org/10.3233/bmr-2008-21108

35. Aytar A, Yuruk ZO, Tuzun EH, Baltaci G, Karatas M, Eker L. The Upper Extremity Functional Index (UEFI): cross-cultural adaptation, reliability, and validity of the Turkish version. J Back Musculoskelet Rehabil. 2015;28(3):489–95. https://doi.org/10.3233/BMR-140545 PMID: 25322741

36. Celik D, Atalar AC, Demirhan M, Dirican A. Translation, cultural adaptation, validity and reliability of the Turkish ASES questionnaire. Knee Surg Sports Traumatol Arthrosc. 2013;21(9):2184–9. https://doi.org/10.1007/s00167-012-2183-3 PMID: 22932692

37. Akel BS, Öksüz Ç, Karahan S, Düger T, Kayihan H. Reliability and validity of Milliken Activities of Daily Living Scale (MAS) in measuring activity limitations of a Turkish population. Scand J Occup Ther. 2012;19(4):315–21. https://doi.org/10.3109/11038128.2011.574153 PMID: 21631171

38. Koldas Dogan S, Ay S, Evcik D, Baser O. Adaptation of Turkish version of the questionnaire Quick Disability of the Arm, Shoulder, and Hand (Quick DASH) in patients with carpal tunnel syndrome. Clin Rheumatol. 2011;30(2):185–91. https://doi.org/10.1007/s10067-010-1470-y PMID: 20411289

39. Tuğay U, Tuğay N, Gelecek N, Özkan M. Oxford Shoulder Score: cross-cultural adaptation and validation of the Turkish version. Arch Orthop Trauma Surg. 2011;131(5):687–94. https://doi.org/10.1007/s00402-010-1242-9 PMID: 21191606

40. Bicer A, Ankarali H. Shoulder pain and disability index: A validation study in Turkish women. Singap Med J. 2010;51(11):865–70.

41. Ayhan Ç, Ünal E, Yakut Y. Turkish version of the simple shoulder test: a reliability and validity study. Turk J Physiother Rehabil. 2010;21(2):68–74.

42. Kitis A, Celik E, Aslan UB, Zencir M. DASH questionnaire for the analysis of musculoskeletal symptoms in industry workers: a validity and reliability study. Appl Ergon. 2009;40(2):251–5. https://doi.org/10.1016/j.apergo.2008.04.005 PMID: 18555973

43. Ozsahin M, Akgun K, Aktas I, Kurtais Y. Adaptation of the Shoulder Disability Questionnaire to the Turkish population, its reliability and validity. Int J Rehabil Res. 2008;31(3):241–5. https://doi.org/10.1097/MRR.0b013e3282fb7615 PMID: 18708847

44. Düger T. Reliability and validity of the Turkish version of the disabilities of the arm, shoulder and hand (DASH) questionnaire. Turk J Physiother Rehabil. 2006;17(3):99–107.

45. Çınar-Medeni Ö, Ozengin N, Baltaci G, Duzgun I. Turkish version of the Rotator Cuff Quality of Life questionnaire in rotator cuff-impaired patients. Knee Surg Sports Traumatol Arthrosc. 2015;23(2):591–5. https://doi.org/10.1007/s00167-014-3290-0 PMID: 25209207

46. Gunes T, Erkorkmaz U, Kurnaz R, Bilgic E, Asci M. Rotator cuff-quality of life scale: adaptation to Turkish. Knee Surg Sports Traumatol Arthrosc. 2015;23(2):603–7. https://doi.org/10.1007/s00167-014-3401-y PMID: 25385526

47. Dogu B, Sahin F, Ozmaden A, Yilmaz F, Kuran B. Which questionnaire is more effective for follow-up diagnosed subacromial impingement syndrome? A comparison of the responsiveness of SDQ, SPADI and WORC index. J Back Musculoskelet Rehabil. 2013;26(1):1–7. https://doi.org/10.3233/BMR-2012-0342 PMID: 23411642

48. El O, Bircan C, Gulbahar S, Demiral Y, Sahin E, Baydar M, et al. The reliability and validity of the Turkish version of the Western Ontario Rotator Cuff Index. Rheumatol Int. 2006;26(12):1101–8. https://doi.org/10.1007/s00296-006-0151-2 PMID: 16799776

49. Duruturk N, Tonga E, Gabel CP, Acar M, Tekindal A. Cross-cultural adaptation, reliability and validity of the Turkish version of the Lower Limb Functional Index. Disabil Rehabil. 2015;37(26):2439–44. https://doi.org/10.3109/09638288.2015.1024342 PMID: 25777549

50. Kocamaz D, Ekici G, Karabulut E. Validation and reliability of the Munich Shoulder Questionnaire for Turkish patients with shoulder dysfunction. Physiother. Quat. 2020;30(1):18–23.

51. Gülcü A, Dirlik G, Çelik D, Kurt EE. The Validation and Verification of the Nottingham Clavicle Score in the Turkish Population. Acta Medica Alanya. 2023;7(1):89–97. https://doi.org/10.30565/medalanya.1287800

52. Şahinoğlu E, Ergin G, Ünver B. Psychometric properties of patient-reported outcome questionnaires for patients with musculoskeletal disorders of the shoulder. Knee Surg Sports Traumatol Arthrosc. 2019;27(10):3188–202. https://doi.org/10.1007/s00167-019-05369-7 PMID: 30739129

53. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol. 1993;46(12):1417–32. https://doi.org/10.1016/0895-4356(93)90142-n PMID: 8263569

54. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25(24):3186–91. https://doi.org/10.1097/00007632-200012150-00014 PMID: 11124735

55. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737–45. https://doi.org/10.1016/j.jclinepi.2010.02.006 PMID: 20494804

56. Wang Y, Yin M, Zhu S, Chen X, Zhou H, Qian W. Patient-reported outcome measures used in patients undergoing total knee arthroplasty. Bone Joint Res. 2021;10(3):203–17. https://doi.org/10.1302/2046-3758.103.BJR-2020-0268.R1 PMID: 33734821

57. Mustafa RA, Santesso N, Brozek J, Akl EA, Walter SD, Norman G, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. J Clin Epidemiol. 2013;66(7):736–42; quiz 742.e1-5. https://doi.org/10.1016/j.jclinepi.2013.02.004 PMID: 23623694

58. Tuğay BU, Tuğay N, Güney H, Hazar Z, Yüksel İ, Atilla B. Cross-cultural adaptation and validation of the Turkish version of Oxford hip score. Arch Orthop Trauma Surg. 2015;135(6):879–89. https://doi.org/10.1007/s00402-015-2215-9 PMID: 25861765

59. Puga VOO, Lopes AD, Costa LOP. Assessment of cross-cultural adaptations and measurement properties of self-report outcome measures relevant to shoulder disability in Portuguese: a systematic review. Rev Bras Fisioter. 2012;16(2):85–93. PMID: 22588123

60. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean?. Ann Rheum Dis. 2007;66 Suppl 3(Suppl 3):iii40-1. https://doi.org/10.1136/ard.2007.079798 PMID: 17934093

61. Longo UG, Saris D, Poolman RW, Berton A, Denaro V. Instruments to assess patients with rotator cuff pathology: a systematic review of measurement properties. Knee Surg Sports Traumatol Arthrosc. 2012;20(10):1961–70. https://doi.org/10.1007/s00167-011-1827-z PMID: 22183737

62. Aldon-Villegas R, Ridao-Fernández C, Torres-Enamorado D, Chamorro-Moriana G. How to Assess Shoulder Functionality: A Systematic Review of Existing Validated Outcome Measures. Diagnostics (Basel). 2021;11(5):845. https://doi.org/10.3390/diagnostics11050845 PMID: 34066777

63. Valderas JM, Ferrer M, Mendívil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health. 2008;11(4):700–8. https://doi.org/10.1111/j.1524-4733.2007.00309.x PMID: 18194398

64. St-Pierre C, Desmeules F, Dionne CE, Frémont P, MacDermid JC, Roy JS. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. Disabil Rehabil. 2016;38(2):103–22. https://doi.org/10.3109/09638288.2015.1027004 PMID: 25801922