Measurement instruments of pain-related avoidance in chronic pain: A systematic review of psychometric properties

Juliane Traxler[abc*], Elena Gaggini[de], Roxane V. Philips[af], Astrid Warny[ag], Madelon Peters[b], Geert Crombez[h], Johan W.S. Vlaeyen[ab]

[a] Research Group Health Psychology, KU Leuven, Leuven, Belgium

[b] Experimental Health Psychology, Maastricht University, Maastricht, The Netherlands

[c] Institute for Health Services Research in Dermatology and Nursing, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[d] University College Maastricht, Maastricht, The Netherlands

[e] London School of Economics and Political Science, London, United Kingdom

[f] Department of Behavioral and Cognitive Science, University of Luxembourg, Luxembourg

[g] Mass General Brigham, Boston, Massachusetts, United States

[h] Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

Text pages: 22

Words (abstract): 239/250

Words (main body): 6107/6000-8000

Figures: 2

Tables: 7

Supplements: 5

* Corresponding author. *Address*: Department of Psychology, KU Leuven, Tiensestraat 102, Box 3726, 3000 Leuven, Belgium. *Phone*: +32 16 32 81 68; *E-mail*: juliane.traxler@kuleuven.be.

Abstract

Avoiding harmful events is adaptive in the short term but eventually may compromise functioning in daily life. Therefore, assessing pain-related avoidance is important in both pain research and treatment. Despite a variety of available measurement instruments, a systematic analysis of their quality and limitations is lacking. We evaluated the measurement properties of instruments used to assess pain-related avoidance in individuals with chronic pain. A systematic review following COSMIN guidelines was conducted (PROSPERO registration: CRD42020181461), including an electronic search of Cochrane, PsycArticles, PubMed, PubPsych, Scopus, and Web of Science as well as grey literature from inception to January 2024. Eligible studies were English, German, or French publications that explicitly claimed to evaluate one or more psychometric properties of measurement instruments assessing pain-related avoidance in adults with chronic pain. Of 703 screened records, 140 original articles were included, covering 20 self-reported questionnaires, one therapist-reported outcome measure, and one performance-based measure. Based on the current evidence, only the Brazilian Portuguese language version of the Chronic Pain Coping Inventory (CPCI) and the Italian version of the CPCI-42 fulfilled criteria to be recommended for use. While the commonly used Fear-Avoidance Beliefs Questionnaire and Tampa Scale of Kinesiophobia were the most extensively studied, the study quality was mixed. The review further highlights extensive research on internal consistency, reliability, and construct validity but reveals a lack of high-quality evidence on measurement error and criterion validity. This work was supported by funding from the Flemish Government (METH/15/011).

Introduction

Pain avoidance is a common and adaptive response to aversive stimuli. Avoidance has been described as "the most prominent component of pain behavior" ([148], p. 274) and occurs in many different forms, ranging from painful movement, activities and daily mobility over housework and work-related contexts to social life [148]. Despite being adaptive, avoidance frequently becomes persistent, even habitual, beyond the regular recovery period. It then may interfere with daily functioning, may lead to withdrawal from valued activities and social life, and may have a detrimental impact on quality of life [3; 149; 153]. Hence, persistent pain-related avoidance behavior is considered maladaptive [107] and according to the Fear Avoidance Model (FAM [201]) contributes to depression, negative affect and, in the long-term, to the transition to chronic pain.

Avoidance can be broadly defined as the prevention of an aversive experience, including the worsening of that experience (e.g., not lifting objects anymore) [176]. It needs to be distinguished from *escape behaviors*, which terminate exposure to the aversive stimulus or situation (e.g., ending a task prematurely because of pain increase), and from *safety behaviors*, referring to actions performed to prevent or minimize harm while enduring the aversive situation (e.g. continuing a task with a supportive device or person) [150].

A variety of questionnaires (e.g., [58; 205]) and behavioral tasks (e.g., [72; 119]) to assess avoidance behavior have been employed in research. However, these instruments have not yet been systematically evaluated and compared, particularly in the light of the challenges associated with measuring avoidance. One challenge is that avoidance can be subtle and may often occur without conscious awareness, making it difficult for individuals to accurately report on it. Additionally, behavioral tasks designed to elicit or model pain-related avoidance may not adequately mimic natural circumstances and may overlook contextual factors. Notwithstanding, accurately measuring avoidance remains highly relevant both for research aimed at understanding the mechanisms underlying avoidance in chronic pain, and for clinical practice aimed at reversing the disabling effects of pain by specifically targeting individual avoidance behavior [37]. These complexities underscore the need for careful deliberation when selecting measurement instruments for assessing pain-related avoidance behavior. Vlaeyen and Linton [202] called for sound assessment techniques for escape and avoidance behavior. Although new instruments have emerged since (e.g., [22; 72]), no attempts have been made to evaluate them systematically. This systematic review identifies available measurement instruments assessing avoidance in adults with chronic pain, and critically appraises their psychometric properties. The goal is to assist researchers and clinicians in choosing an appropriate measurement instrument of pain-related avoidance.

Methods

A systematic literature search for instruments designed to measure pain-related avoidance behavior was conducted. It was conducted and reported according to a protocol registered and published on the International Prospective Register of Systematic Reviews (PROSPERO; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=181461, registration number: CRD42020181461) prior to the formal screening of the search results. For reasons of feasibility, the inclusion criteria have been further refined, and the study population has been limited to persons with chronic pain after the initial search (all amendments have been registered in PROSPERO). The updated COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN; [126]) guidelines for systematic reviews of patient-reported outcome measures was used and the PRISMA-COSMIN reporting guideline [49] was followed. The COSMIN framework is used for evaluative purposes (for an overview and comparison of this and other frameworks, see [109; 167]).

Search Strategy & Eligibility Criteria

The databases searched included Cochrane, PsycArticles, PubMed, PubPsych, Scopus, and Web of Science, without restriction by publication period. The searches were conducted in April and June 2020 and were re-run prior to the final analysis in January 2024.

The search strategy is described in detail in Appendix A. The search categories *avoidance behavior*, *pain*, and *psychometric properties* were combined, while filtering out animal, pharmacological, and pediatric population studies. Articles were eligible if they (a) reported on the development, or on one or more psychometric properties of at least one measurement instrument of pain-related avoidance behavior; (b) studied an adult human population (> 18 years) with chronic pain, except for headaches, migraine, and cancer pain; (c) were written in English, French or German, and (d) were available in full-text. Exclusion criteria comprised: (a) reviews and meta-analyses; (b) conceptual and narrative papers discussing pain-related avoidance behavior without adding significant information on measurement properties; (c) studies that used the instrument purely as an outcome measure [151]; (d) pharmacological studies. A post hoc decision was made to exclude studies that exclusively focused on avoidance in headache/migraine due to feasibility and the different types of triggers involved in these pain conditions.

Preliminary searches were conducted by JT to establish the need for and timeliness of a review on this topic, and to determine the study selection process, which was piloted subsequently. References and citing references of included studies were manually searched. After duplicate removal, titles and abstracts from the definite database search were screened for their eligibility by two independent reviewers (AW and RVP). In case of disagreement between the reviewers, a third reviewer (JT) was consulted to reach a final decision. Next, the full texts were screened. If full texts could not be retrieved, study authors were contacted (n = 3). An expert panel was consulted to ensure that all relevant instruments were included.

Data Extraction

One reviewer (JT) extracted the following information from the included studies: author(s) and year of publication, study population and demographics, pain characteristics, and characteristics and psychometric properties of measurement instruments. Data were extracted onto a bespoke data extraction table. A second reviewer (CK, EG) cross-checked the accuracy of the extracted information of a random 50% of included records, stratified for each instrument. Disagreements were resolved through discussion. Data extracted from included studies and used for all analyses can be found in Tables 1 and 2 and Appendix E.

Evaluation of Measurement Properties and Risk of Bias Assessment

The measurement properties of self-report instruments were evaluated according to the guidelines by [151] for systematic reviews of patient-reported outcome measures (PROMs) consisting of three steps: first, the methodological quality of individual studies was appraised for each measurement separately based on the COSMIN Risk of Bias checklist [1] using a four-point rating system (very good, adequate, doubtful, inadequate). The "worst score counts" principle was applied [187]. Next, the psychometric properties of measurements instruments reported in each included study were rated as sufficient (+), insufficient (-) or indeterminate (?) using the COSMIN updated criteria for good measurement properties ([186], see Table 3). These steps were first applied by two independent reviewers (JT, CK) to evaluate content validity. In case of disagreements, consensus was reached through discussion. Subsequently, following the steps described, one reviewer (JT) assessed the internal structure of the PROMs (i.e., structural validity, internal consistency, cross-cultural validity) – for which a reflective measurement model was assumed, unless otherwise specified in the instrument development paper - and the remaining measurement properties test-retest reliability, measurement error, criterion validity, hypothesis testing for construct validity, and responsiveness to change [125]. Twenty percent of these assessments were cross-checked by an independent reviewer (EG).

Criterion validity is tested by comparing an instrument to a criterion measure that is known to accurately assess the construct of interest. According to COSMIN, the gold standard should be used but this does not exist for PROMs, except if a shortened version is compared to the original instrument. Comparisons to, for instance, other widely used instruments are considered hypothesis testing for construct validity [125]. Given both the variety in the included PROMs and the large variability in study designs, we did not define global a priori hypotheses against which all study results would be compared to evaluate construct validity and responsiveness to change but instead followed the approach of [76]. We extracted hypotheses relating to these measurement properties from the articles or, if the authors had not specified a priori hypotheses, we deduced hypotheses from the authors' descriptions (Appendix B and C). This approach was slightly modified for use with performance-based outcomes measures [77].

[insert Table 3 here]

Data Synthesis

Evidence from multiple studies on the same measurement instrument was summarized per measurement property. A qualitative synthesis of the evidence per psychometric property was performed for each measurement instrument by one reviewer, resulting in a rating of sufficient (+), insufficient (-), inconsistent (±) or indeterminate (?), and graded following the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach: quality of evidence was down-graded due to risk of bias, inconsistency, imprecision, and/or indirectness, resulting in "high", "moderate", "low" or "very low" quality [151]. When results of studies on the same measurement instrument were inconsistent, sources for inconsistency such as differences in populations were explored. When inconsistency could not be explained, results were summarized based on the consistent majority findings, if possible, or rated as inconsistent (±) without summarizing. Based on the evaluation of measurement properties, recommendations for the use of instruments assessing pain-related avoidance behavior were formulated according to the guidelines by [151] by sorting each PROM into one of three categories:

A: PROM can be recommended for use [evidence for sufficient content validity (any level), and at least low-quality evidence for sufficient structural validity and internal consistency];

B: PROM has potential to be recommended for use, but further validation is needed (PROM cannot be categorized into A or C);

C: PROM should not be recommended for use (high-quality evidence for insufficient measurement criteria).

Results

The database search and the manual search identified 658 and 182 articles, respectively. After removing 170 duplicates, 669 records were screened, of which 303 were excluded and 366 records were retrieved. Of these, 227 records were ineligible (for a list of excluded records with reasons for exclusion according to the inclusion criteria, see Appendix D): 5 records could not be retrieved, 151 did not report data on psychometric properties of relevant instruments, 24 were not original articles, 27 addressed a different target group, 13 did not address avoidance, and 6 were published in a language other than English, French, or German. A flow chart illustrating the selection of articles is presented in Figure 1.

The 140 eligible articles comprised 20 different PROMs [Table 1; Avoidance Daily Activities Photo Scale (ADAP; n = 2); Avoidance-Endurance Questionnaire (AEQ; n = 3); Activity Patterns Scale (APS; n = 1); Burn Survivor Fear-Avoidance Questionnaire (BSFAQ;

n = 1); Catastrophizing Avoidance Scale D-65+ (CAS-D-65+; n = 1); Chronic Pain Coping Inventory (CPCI; n = 7); Fear-Avoidance Beliefs Questionnaire (FABQ; n = 50); Fear-Avoidance Components Scale (FACS; n = 9); Negative Responsivity to Pain Scale (NRP; n = 1); Pain and Activity Relations Questionnaire (PARQ; n = 1); Pain Anxiety Symptoms Scale (PASS; n = 2); Pain-Coping Inventory (PCI; n = 1); Psychological Inflexibility in Pain Scale (PIPS; n = 7); Patterns of Activity Measure - Pain (POAM-P; n = 2); Tampa Scale of Kinesiophobia (TSK; n = 32); Tampa Scale of Kinesiophobia-Temporomandibular Disorders (TSK-TMD; n = 4)], four abbreviated versions of original PROMs [Table 1; CPCI-42 (n = 6); PASS-20 (n = 5); TSK-11 (n = 15); TSK-13 (n = 5)], and two behavioral measures [Table 2; Behavioral Avoidance Test-Back (BAT-Back; n = 2); clinician-reported fear avoidance beliefs (n = 1)], with some studies investigating more than one instrument. The studies investigated content validity (n = 37), structural validity (n = 59), internal consistency (n = 99), test-retest reliability (n = 85), measurement error (n = 27), criterion validity (n = 4), construct validity (hypothesis testing; n = 113), and responsiveness to change (n = 24). In the following, results will be described only for the original language version of each instrument and notable differences with other language versions will be pointed out. Characteristics of the included measurement instruments are displayed in Tables 1 and 2, respectively. Individual study characteristics are presented in Table 4. The risk of bias assessments and rating against measurement properties of included studies can be found in Tables 5 (content validity) and 6 (all other measurement properties); the evidence synthesis including rating of results and overall level of evidence are shown in Table 7. Figure 2 shows the number of studies investigating each of the measurement properties, including the risk of bias (RoB) ratings.

[please insert Figure 1 here]

[please insert Figure 2 here]

[please insert Table 1 here]

[please insert Table 2 here]

[please insert Table 4 here]

Content validity

Quality of PROM development studies

Table 5 provides an overview of the overall results and quality of the 16 available development studies and the content validity studies. As no patients were involved in the development process, concept elicitation was judged to be inadequate for all PROMs, except for the Avoidance Daily Activities Photo scale (ADAP) [8]. The concept elicitation of the ADAP was described superficially and resulted in a "doubtful" rating. Furthermore, only the development studies of APS, FACS, PIPS and TSK-TMD included cognitive interviews with patients. The process for PIPS was deemed inadequate because too few patients were involved. The quality of cognitive interviews for APS and TSK-TMD was doubtful since relevant information was not reported in the papers. For instance, information on whether interviews were based on appropriate guides and recorded and if at least two researchers were involved in the analysis was not presented. The cognitive interview for FACS was of doubtful quality as it was purely based on a quantitative approach. No development studies could be retrieved for BSFAQ, PCI, TSK, or TSK-13.

Quality and results of content validity studies

The 37 articles on content validity comprised 35 studies involving patients, of which all assessed comprehensibility, five additionally addressed relevance, and none examined comprehensiveness, and ten studies involving professionals, of which nine assessed relevance and five assessed comprehensiveness. Thirty-two were cross-cultural adaptations that pretested the translated PROM. One study [135] was of inadequate quality as the method asking patients about the relevance of the FACS was inappropriate; the quality of all other studies was deemed doubtful as relevant information was not reported. No content validity studies on AEQ, BSFAQ, CAS-D-65+, NRP, PARQ, PASS, PCI, and TSK-13 were found.

Evidence synthesis

Overall, no high-quality evidence on content validity was available for any of the instruments. For CAS-D-65+, the quality of evidence for indeterminate results was “moderate” based on the reviewers’ ratings and the development study. For TSK, the quality of evidence for sufficient content validity was “moderate” based on the reviewers’ ratings and content validity studies.

There was “low” evidence for sufficient relevance for ADAP, CPCI, CPCI-42, PCI, and TSK-TMD. The quality of evidence for relevance was “very low” for all other instruments, often because the rating was solely based on reviewers’ ratings [26]. Regarding comprehensiveness, quality of evidence was “moderate” for FABQ and TSK with sufficient results based on the reviewers’ ratings and content validity studies. For all other instruments, quality of evidence for comprehensiveness was “very low”. Lastly, there was “moderate” quality evidence for sufficient comprehensibility for FABQ, FACS, PASS-20, PAOM-P, TSK, TSK-11, and TSK-TMD, “low” quality evidence for sufficient comprehensibility for ADAP, APS, CPCI, and CPCI-42, and “very low” quality evidence for sufficient comprehensibility for AEQ, PARQ, PASS, and PCI. “Very low” evidence pointed to inconsistent results for CAS-D-65+ and NRP, and to indeterminate results for PIPS.

As no high-quality evidence supporting insufficient content validity was found for any of the PROMs, the other psychometric properties were assessed.

[please insert Table 5 here]

Structural validity

Eighty studies assessed structural validity of the included PROMs in persons with chronic pain. Most of the studies (40%) were of very good quality, 41.25% were of adequate quality, 5% of doubtful quality, and 13.75% of inadequate quality. Inadequate ratings were mostly due to small sample sizes.

Sufficient structural validity was found for APS, NRP, PIPS (all language versions except Greek), TSK-TMD, the Brazilian Portuguese and French versions of CPCI, the English version of CPCI-42, the Dutch and German versions of TSK-11, and the Serbian version of FACS (all high evidence), the Dutch version of the TSK-13 and the Italian version of the CPCI-42 (both moderate evidence). There was high quality evidence for insufficient structural validity for the PASS-20, TSK (English, Dutch), TSK-11 (English, Swedish), TSK-13, and for the Swiss-German and Chinese versions of the FABQ. Lastly, structural validity was indeterminate for TSK (high quality evidence), ADAP, FABQ, FACS and PARQ (moderate evidence), and AEQ and PCI (low to very low evidence). No studies were retrieved that tested structural validity of BSFAQ, CAS-D-65+, PASS, PCI, or POAM-P.

Internal consistency

One hundred-eleven studies assessed internal consistency, covering all the included PROMs, except for BSFAQ. The vast majority of studies (90.09%) were of very good quality, while the quality of 11 studies (9.91%) was rated as inadequate because they did not calculate an internal consistency statistic for each unidimensional (sub)scale separately.

Internal consistency was sufficient for CPCI-42, NRP, PIPS, TSK-TMD and the Dutch version of the TSK-13 (all high evidence). High quality evidence was found for insufficient internal consistency for the Dutch and German versions of TSK-11 and the Dutch and Italian versions of TSK. For APS, internal consistency was considered inconsistent as the two available high-quality studies in one publication [51] found contrasting results, despite comparable samples. For all other PROMs, internal consistency was rated indeterminate as the prerequisite of at least low-quality evidence for sufficient structural validity was not met, even if Cronbach's alpha > 0.7.

Reliability

Eighty-five studies assessed test-retest reliability of most PROMs. There is no information on test-retest reliability available for the APS, BSFAQ, PARQ and PASS. Only two studies (2.35%) were of very good quality, 16.47% were of adequate quality while 63.53% were of doubtful and 17.65% of inadequate quality. Poor quality ratings were mostly due to inappropriate time intervals and uncertainty about whether patients were stable in the interim period and/or whether test conditions were similar at both measurement time points.

Reliability was deemed sufficient for TSK-11 (high evidence), CPCI-42, FABQ, TSK (all moderate evidence), ADAP, PASS-20, POAM-P (French version; all low evidence), and CAS-D-65+ (very low evidence). For CPCI, FACS, NRP, PCI, and TSK-TMD, test-retest reliability was considered indeterminate.

Measurement error

Twenty-seven studies assessed measurement error of ADAP, AEQ, FABQ, FACS, TSK, TSK-11, TSK-13 and TSK-TMD. Of these, two studies each (7.41%) were of very good [48; 127] and adequate [16; 81] quality, 70.37% were of doubtful quality and the remaining studies (14.81%) were of inadequate quality. Again, poor quality ratings can be ascribed to uncertainty about or dissimilarity between test conditions for the measurements in all of these studies, and the uncertainty about whether patients were stable in the interim period.

One study provided high [127] and low [128] quality evidence, respectively, for sufficient measurement error for the Italian version of the TSK. Two studies provided low evidence for insufficient measurement error for the Igbo version of the FABQ [74] and the Dutch version of the TSK [155]. In all other cases, measurement error was rated as indeterminate because none of the studies defined the minimal important change.

Criterion validity

Four studies assessed criterion validity of CPCI-42 [166], FABQ [80], PASS-20 [113] and TSK-11 [174]. Evidence for FABQ was indeterminate, while high quality evidence for sufficient criterion validity was found for the other three instruments.

Hypothesis testing for construct validity

Construct validity via hypothesis testing (convergent validity) was assessed in 111 studies for all instruments. 14.41% were of very good quality, 26.13% were of adequate quality, 13.51% were of doubtful quality, and 45.95% of studies were of inadequate quality. Known-groups validity was assessed in 18 studies for AEQ, BSFAQ, FABQ, FACS, PCI, TSK, and TSK-11. 33.33% of studies were of very good quality, 16.67% were of adequate quality, 38.89% were of doubtful quality, and 11.11% were of inadequate quality. Insufficient description of important characteristics of the subgroups accounted for most of the low-quality ratings (doubtful, inadequate).

Convergent validity and known-groups validity coupled together, there was high quality evidence for sufficient construct validity for the English language versions of CPCI and TSK. For most other instruments, more than 75% of hypotheses were confirmed as well but at lower quality evidence levels. Construct validity was insufficient with moderate to high-quality evidence for FABQ (Greek), TSK (Gujarati), and TSK-11 (Brazilian Portuguese, Chinese, Spanish), and with low to very low-quality evidence for ADAP, CPCI-42 (Korean), FABQ (Dutch, Gujarati, Hausa, Igbo, Marathi, Swedish, Swiss-German, and Yoruba language versions), PASS (Dutch), and TSK-13. It was inconsistent for TSK-11 (moderate evidence). Lastly, construct validity was indeterminate for FABQ (Arab, English, and Hindi language versions), and TSK-TMD (Chinese and Dutch language version).

Responsiveness to change

Responsiveness to change was assessed in 25 studies for CPCI-42, FABQ, FACS, POAM-P, TSK, and TSK-11, some of which used more than one approach. 28.21% of studies were of very good quality, 10.26% were of adequate quality, 30.77% were of doubtful quality, and 30.77% were of inadequate quality.

Responsiveness to change was deemed sufficient for CPCI and CPCI-42 (very low evidence), and FABQ, TSK (English and Italian versions) and TSK-11 (high quality evidence). However, results were not in line with at least 70% of a priori hypotheses for other language versions of the FABQ (Brazilian Portuguese, Chinese, Greek, Turkish), FACS, the Dutch version of the TSK, TSK-11 (Brazilian Portuguese), and TSK-13 (Portuguese version), indicating low responsiveness to change. Lastly, there was very low evidence for indeterminate responsiveness to change for POAM-P.

[please insert Table 6 here]

[please insert Table 7 here]

Recommendations

Only the Brazilian Portuguese language version of the CPCI and the Italian version of the CPCI-42 fulfilled all criteria of category A to be recommended for use. Importantly, these scales do not include an avoidance subscale but two subscales called "resting" and "guarding", which were considered here. Based on the current evidence, the Chinese, Gujarati and Swiss-German versions of the FABQ, the Chinese version of the PASS-20, the Greek version of the PIPS, the Chinese, Dutch and Italian versions of the TSK, the English, Dutch, German and Swedish versions of the TSK-11 and the English version of the TSK-13 cannot be recommended due to high-quality evidence for insufficient structural validity and/or internal consistency (category C). All other instruments and language versions have the potential to be recommended for use, pending further validation (category B): While several language versions of APS, CPCI-42, FACS, NRP, PIPS, TSK-13 and TSK-TMD performed well in terms of most psychometric properties, their content validity could not clearly be classified as sufficient and requires further assessment.

Non-validated measures of avoidance behavior

In addition to the self-report instruments described above, a plethora of instruments are used to measure pain-related avoidance in research that have not been formally validated according to classic test theory, and thus did not meet the inclusion criteria of this review. These comprise predominantly behavioral tasks, most of which operationalize avoidance in terms of response rate [11; 82; 195] or repetitions of certain movements [147]. Other instruments assess avoidance by measuring the willingness to repeat a painful task, such as the cold-pressor test [28] or an ischemic pain task [41]. Moreover, response latency (e.g., Voluntary Joystick Movement paradigm [120]), overall task duration [82], range of motion [147] or the deviation from a painful movement as used in the HapticMaster paradigm [119] are means of measuring avoidance. Lastly, rather than avoidance of painful movements or experimental pain stimuli, the Approach Avoidance Task [137] is based on pain-related images that can be avoided by zooming out. Although none of these tasks have been psychometrically validated yet, evidence of construct validity is often reported, and they bear great potential as most of them do not rely on self-report and have a comparatively high ecological validity due to their use of actual movements. Notwithstanding, many of these paradigms bear different limitations. As stated by Meulders et al. [118], the tasks rely on dichotomous responses or examine adaptive and/or low-cost avoidance which may ultimately limit their utility.

Proxy measures of pain-related avoidance

The expert panel suggested several instruments for potential inclusion in this review, including the Acceptance and Action Questionnaire II – Pain Version (AAQ-II-P) [159], the Coping Strategies Questionnaire (CSQ) [168], the Chronic Pain Acceptance Questionnaire (CPAQ) and its revised version (CPAQ-R) [115], the Pain and Impairment Relationship Scale (PAIRS) [160], and the Pain Interference Index (PII) [85]. However, after careful deliberation, these instruments were excluded as they primarily measure related yet distinct constructs: experiential avoidance (AAQ-II-P), acceptance (CPAQ, CPAQ-R), coping (CSQ), or interference (PII; assessing the extent to which pain has made certain activities difficult) rather than avoidance behavior per se. The PAIRS includes some items that could arguably reflect aspects of avoidance; however, to our knowledge, it has not been specifically utilized for this purpose. Additionally, apart from the AAQ-II-P and the CSQ, none of these instruments were identified in the literature search.

Several other instruments used as proxy measures should be highlighted as well: The Photograph Series of Daily Activities (PHODA; [98]) and its shortened electronic version (PHODA-SeV; [108]) primarily examine the perceived harmfulness of daily activities and movements as shown in a set of photographs. The German scale Ältere Menschen in körperlicher Aktion (eng.: "Elderly people in physical action", AMIKA [156]) and its short 8-

item version, AMIKA-K [158], are based on the PHODA but adapted for use in people aged 65 and older with low back pain. Similarly, the Pictorial Fear of Activity Scale-Cervical (PFActS-C; [193]) uses photographs of various activities that specifically strain the neck to assess a person's worry or fearfulness about carrying out the displayed activities.

While these measurement instruments are occasionally used as proxy measures for avoidance and the instructions could be adjusted for this purpose, they do not explicitly target avoidance behavior. To maintain a focused and feasible scope, the review was narrowed to instruments specifically designed to measure avoidance.

Discussion

The purpose of this review was to systematically evaluate and compare the quality of instruments used to measure pain-related avoidance behavior, with the aim of providing recommendations for their use in evaluative clinical and research settings. Given that measurement properties may greatly affect study results, the use of high-quality instruments is pivotal to ensure reliable conclusions in both research and treatment contexts. To achieve trustworthy results, this review followed the COSMIN guidelines for systematic reviews. While two recent COSMIN reviews have been conducted on the measurement properties of specific instruments assessing pain-related avoidance – the TSK [47] and the cross-culturally adapted versions of the TSK-11 [7] – this is, to our knowledge, the first comprehensive COSMIN review of all available validated instruments addressing pain-related avoidance.

We identified 21 versions of 16 unique instruments. Of these, only the Brazilian Portuguese language version of the Chronic Pain Coping Inventory (CPCI) [179; 180] and the Italian version of the CPCI-42 [130] show sufficient content validity, structural validity, and internal consistency [151] and can be recommended for use based on the evidence currently available. However, it should be noted that these recommendations are based on just two and one publication, respectively. Although according to the COSMIN guidelines, a single high-quality study is sufficient to rate psychometric properties, a low number of publications means a lower chance for inconsistent findings compared to well-studied instruments like the FABQ or TSK. While some of the other instruments show insufficient structural validity and/or internal consistency and, therefore, cannot be recommended for use, most instruments require further validation with respect to their content and structural validity before they can be recommended or advised against.

As content validity is considered the most important measurement property [152], the lack of adequate evidence across instruments is concerning. In fact, content validity has been examined in most PROMs to some degree; however, most development studies did not include people with lived experience (PWLE) of pain in PROM development or the concept elicitation. Many PROM developments did not include a cognitive interview. The relevance, comprehensibility and especially comprehensiveness were often not investigated. It is, therefore, highly recommended to adhere to modern standards and include PWLE for the development of new instruments, to continue testing relevance, comprehensiveness, and comprehensibility in existing PROMs, and to update and improve instruments, if necessary.

Apart from content validity, the most frequently reported psychometric properties include internal consistency, reliability, and construct validity, whereas measurement error, responsiveness to change and criterion validity – according to the gold standard approach by COSMIN – received little attention. Importantly, quality of evidence was mixed, largely due to shortcomings in reporting, e.g. a lack of information about whether patients were stable in the interim period and/or whether test conditions were similar at both measurement time points (test-retest reliability, measurement error), or an insufficient description of subgroup characteristics (known-groups validity) or of the intervention used in responsiveness to change studies.

We would like to highlight some general challenges associated with measuring pain-related avoidance behavior, which may affect the validity and reliability of measurement instruments. First, avoidance is not necessarily a conscious behavior and can be subtle, so that people may lack insight. Second, an additional common problem of self-report is recall bias, as people may fail to correctly reproduce the degree of and circumstances under which they engage in avoidance behavior. Third, patients may be reluctant to report on avoidance behaviors for fear of stigmatization and instead give socially desirable answers. These problems complicate the use of self-report instruments to reliably measure avoidance behavior [204]. Most behavioral tasks, on the other hand, do not adequately reflect everyday life. They usually constitute a comparatively safe, predictable, and controllable context, often employ a simple operationalization of avoidance [95; 118] and tend to disregard competing motivations [202]. Research in people living with chronic pain [50; 84; 143] and pain-free participants [29-31; 195] demonstrated that goals and competing motivations can attenuate avoidance, even when pain-related fear remains high. People living with chronic pain may have reasons to avoid a movement, posture or activity in their daily life but not in the clinic, and vice versa. For example, holding an arm overhead to hold on to a grab handle on the bus that may stop abruptly is more consequential and more likely to be perceived as threatening than performing the same movement in isolation on a stable office floor. Conversely, a patient may be reluctant to lift a heavy object as part of a clinical assessment, which bears low intrinsic value, but not hesitate to pick up their grandchild. Furthermore, avoidance may be influenced by mood states [83; 203] and, thus, fluctuate over time [205]. The dynamics of avoidance behavior are also well demonstrated from the perspective of the exploitation-exploration trade-off [96] and Pavlovian-Instrumental Transfer literature [97]. These findings suggest that assessment of avoidance behavior in one specific context or even in a decontextualized way using one of the recommended questionnaires may not capture an accurate picture of a person's pain avoidance behavior. Alternatives to these assessments are observations of patients' avoidance behavior in everyday life. As this is usually not feasible for clinicians and researchers alike, avoidance can be measured in real life using ecological momentary assessment (EMA; see e.g. [178]). The advantages of this methodology are that a person's behavior is assessed repeatedly in real-time and in the person's natural environment, thus minimizing recall bias [57]. Hence, EMA offers a highly flexible and ecologically valid approach, although the challenges of lack of insight and potential reluctance to report avoidance remain. Considering the shortcomings of each approach, we recommend keeping these limitations of even well-validated self-report instruments in mind and basing clinical decisions not on a single measurement but using different types of assessment and sources of information where possible. One should also keep in mind that self-reports may only provide valid information in the context of a non-stigmatizing, trusting and empathic relationship.

Several limitations related to both the evidence included in this review and the review process warrant attention. First, several of the included studies were of questionable or insufficient quality, often due to poor or unclear reporting, resulting in low-quality evidence or inconclusive findings. Importantly, although not all studies specified the medium through which the PROMs were completed (e.g. paper-based vs. digital formats), it is likely that most were administered in a paper-based format. Therefore, future research should also assess the validity and reliability of computerized versions of these instruments [108]. Second, aside from the FABQ, the TSK and the TSK-11, most measurement instruments have been translated and culturally adapted into only few or no non-Western languages. This lack of translation limits research and clinical attention to avoidance behaviors in diverse populations across the globe. Third, we noted that psychometric terminology is often used incorrectly in the literature, which may be misleading for readers. Consequently, we recommend further research into the psychometric properties of measurement instruments of pain-related avoidance using terminology and frameworks consistently and encourage efforts to translate and validate more instruments in other languages, in particular, in African and Asian languages. As a side note, the included studies almost exclusively used Cronbach's alpha to measure internal consistency. Although this is the most common approach, the use of this measure has been criticized because the assumptions are often not met, and alternatives such as McDonald's omega have been proposed [117].

Regarding the review process itself, we only included studies that specifically aimed at testing one or multiple instruments' psychometric properties. Many other empirical studies provide information on internal consistency and construct validity. As reviewing and integrating all these records was not possible, this review should be considered a start that can be built upon. Similarly, for feasibility reasons, only one reviewer conducted the risk of bias and measurement property assessments, with an independent cross-check of a subset of publications by one of two researchers. While no frequent or systematic mistakes were identified in the randomly selected subset, we acknowledge that a small risk for error may remain. Moreover, we used the "worst score counts" method, adhering to the COSMIN guidelines. This may not sufficiently distinguish between subtle differences in the methodological quality of the included studies [181]. As highlighted above, the number of publications available for each individual instrument and language version may have influenced the recommendations for use. Both recommended instruments as well as most of those that could explicitly not be recommended were supported by evidence from only one or two studies. The instruments categorized as needing further validation were often more extensively studied, with inconsistent evidence. Hence, we strongly advise further investigation of the Brazilian Portuguese version of the CPCI and the Italian version of the CPCI-42. In addition, we evaluated the instruments for their evaluative purpose, and we assumed that measures were designed using a reflective measurement model. The evaluation of the psychometric properties in this paper reflects these assumptions. However, other purposes and measurement models are possible, requiring different psychometric standards [196]. For instance, good responsiveness to change is critical if an instrument is used to evaluate changes in avoidance behavior in response to an intervention but may be unnecessary in the context of diagnosis and screening, where known-groups validity and the availability of cut-off scores play an important role instead [37; 39; 196].

Future studies may extend this work with records published in languages other than English, French or German. Furthermore, we recommend systematic reviews on measurement instruments for use in (1) pain-free research participants and (2) children and adolescents experiencing chronic pain.

The clinical implications of this review are two-fold. Firstly, even though only two instruments can be recommended based on the current evidence, this overview of measurement properties of available instruments enables researchers and clinicians to select an instrument based on their specific goals. For example, even if not all measurement properties are well-tested, an instrument with good content validity, internal consistency, and responsiveness to change may be a viable option for evaluating treatment outcomes. Secondly, as stated before, we highly recommend complementing standardized questionnaires with EMA and/or observations, for example using movement tasks like the BAT-Back or through report by family members, to capture a person's behavior in different contexts. Importantly, to facilitate interpretation of assessment results and inform clinical decision making, norms and cut-off scores are needed (for example using regression modelling, e.g. [194]). These are available for only few of the instruments reviewed, including the FABQ [165] and TSK [54], and were established in specific languages and populations, limiting their generalizability. Hence, more work on norm scores and cut-offs is needed.

This review is the first to systematically capture all instruments commonly used to measure avoidance behavior in adult patients with chronic pain and to comprehensively evaluate their psychometric properties in accordance with the COSMIN guidelines. We hope this overview and the recommendations may aid researchers and clinicians in choosing the best-suited instruments. Furthermore, our findings reveal a notable lack of high-quality evidence supporting key measurement properties, even among the most frequently utilized PROMs. Addressing this critical gap through robust validation studies is essential to ensure the reliability and validity of clinical and research findings that rely on these measures.

Acknowledgements

The authors wish to thank the expert panel consisting of Christopher Eccleston, Monika Hasenbring, Mark Jensen, Lance McCracken, Laura Simons, Rikard Wicksell and Amanda C. de C. Williams for their valuable input. Furthermore, the authors thank Céline Kleckner for her help with data extraction and risk of bias rating. This work was supported by the "Asthenes" long-term structural funding Methusalem grant (METH/15/011) from the Flemish Government, Belgium. The authors have no conflicts of interest to declare. All authors contributed to and approved the final version submitted.

References

[1] COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) website. http://www.cosmin.nl.

[2] Abad SKG, Akhbari B, Salavati M, Saeedi A, Seydi M, Shakoorianfard MA. Translation, reliability, and validity of the avoidance endurance questionnaire in Iranian subjects with chronic non-specific neck pain. Journal of Family Medicine and Primary Care 2020;9(7):3565.

[3] Abbott AD, Tyni-Lenné R, Hedlund R. The influence of psychological factors on pre-operative levels of pain intensity, disability and health-related quality of life in lumbar spinal fusion surgery patients. Physiotherapy 2010;96(3):213-221.

[4] Aguiar AdS, Bataglion C, Visscher C, Bevilaqua Grossi D, Chaves TC. Cross-cultural adaptation, reliability and construct validity of the Tampa scale for kinesiophobia for temporomandibular disorders (TSK/TMD-Br) into Brazilian Portuguese. Journal of oral rehabilitation 2017;44(7):500-510.

[5] Alanazi F, Gleeson P, Olson S, Roddey T. Translation and validation of the Arabic version of the Fear-Avoidance Beliefs Questionnaire in patients with low back pain. Spine 2017;42(7):E411-E416.

[6] Alanazi FS. Cross-Cultural Adaptation and Validation of the Arabic Version of the Tampa Scale of Kinesiophobia in Patients with Low Back Pain. Hail Journal of Health Sciences 2021;3(1):34-46.

[7] Alpalhão V, Vaz JR, Cordeiro N, Pezarat-Correia P. Are the cross-culturally adapted versions of the Tampa Scale for Kinesiophobia 11-item valid, reliable, and responsive? A COSMIN-informed systematic review of measurement properties. The Journal of Pain 2024:104602.

[8] Ansanello W, Dos Reis FJJ, Tozzo MC, Zatiti SCA, Meulders A, Vlaeyen JW, de Oliveira AS. Development of the avoidance daily activities photo scale for patients with shoulder pain. Physical Therapy 2022;102(2):pzab268.

[9] Ansanello W, Dos Reis FJJ, Tozzo MC, Zatiti SCA, Meulders A, Vlaeyen JW, de Oliveira AS. Reliability and Validity of the Avoidance of Daily Activities Photo Scale for Patients With Shoulder Pain (ADAP Shoulder Scale). Physical Therapy 2023;103(12):pzad101.

[10] Areeudomwong P, Buttagat V. Reliability and validity of the cross-culturally adapted thai version of the tampa scale for kinesiophobia in knee osteoarthritis patients. The Malaysian journal of medical sciences: MJMS 2017;24(2):61.

[11] Arntz A, Van Eck M, de Jong P, Van den Hout MA. The relationship between underpredicted pain and escape. Behaviour research and therapy 1990;28(1):87-90.

[12] Askary-Ashtiani A, Ebrahimi-Takamejani I, Torkaman G, Amiri M, Mousavi SJ. Reliability and validity of the persian versions of the fear avoidance beliefs questionnaire and tampa scale of kinesiophobia in patients with neck pain. Spine 2014;39(18):E1095-E1102.

[13] Barke A, Riecke J, Rief W, Glombiewski JA. The Psychological Inflexibility in Pain Scale (PIPS)–validation, factor structure and comparison to the Chronic Pain Acceptance Questionnaire (CPAQ) and other validated measures in German chronic back pain patients. BMC Musculoskeletal Disorders 2015;16:1-10.

[14] Benaim C, Léger B, Vuistiner P, Luthi F. Validation of the French version of the "Patterns of Activity Measure" in patients with chronic musculoskeletal pain. Pain Research and Management 2017;2017(1):6570394.

[15] Bid DD, Alagappan TR. Construct validity and factor analysis of the Gujarati version of the fear-avoidance beliefs questionnaire. Physiotherapy-The Journal of Indian Association of Physiotherapists 2019;13(1):30-37.

[16] Bid DD, Neblett R, Alagappan TR, Patel CJ, Patel KN, Patel RL, Narola SJ, Sailor VV. Cross-cultural adaptation, reliability, and validity of the Gujarati fear-avoidance components scale. Physiotherapy-The Journal of Indian Association of Physiotherapists 2020;14(2):98-107.

[17] Bid DD, Ramalingam AT, Sinha SR, Rathi PB, Patel VN, Rajwani JM, Patel KN. Cross-cultural adaptation, reliability, validity, and factor analysis of the Gujarati version of the Tampa scale of kinesiophobia in chronic low back pain. Physiotherapy-The Journal of Indian Association of Physiotherapists 2018;12(2):79-86.

[18] Bid DD, Soni NC, Rathod PV, Ramalingam AT, Sinha RK. Cross cultural adaptation, reliability and validity of Gujarati version of fear avoidance belief questionnaire in chronic low back pain. Natl J Integr Res Med 2016;7:1-8.

[19] Burwinkle T, Robinson JP, Turk DC. Fear of movement: factor structure of the Tampa Scale of Kinesiophobia in patients with fibromyalgia syndrome. The Journal of pain 2005;6(6):384-391.

[20] Cai L, Liu Y, Woby SR, Genoosha N, Cui M, Guo L. Cross-cultural adaptation, reliability, and validity of the Chinese version of the Tampa Scale for Kinesiophobia-11 among patients who have undergone total knee arthroplasty. The Journal of arthroplasty 2019;34(6):1116-1121.

[21] Calley DQ, Jackson S, Collins H, George SZ. Identifying patient fear-avoidance beliefs by physical therapists managing patients with low back pain. Journal of orthopaedic & sports physical therapy 2010;40(12):774-783.

[22] Cane D, Nielson WR, McCarthy M, Mazmanian D. Pain-related activity patterns: measurement, interrelationships, and associations with psychosocial functioning. The Clinical journal of pain 2013;29(5):435-442.

[23] Chaory K, Fayad F, Rannou F, Lefèvre-Colau M-M, Fermanian J, Revel M, Poiraudeau S. Validation of the French version of the fear avoidance belief questionnaire. Spine 2004;29(8):908-913.

[24] Chen J, Caluori C, Alberton L, Zhang J, Shashoua D, Calva V, Gauthier N, Edger-Lacoursière Z, de Oliveira A, Marois-Pagé E. Validation of the burn survivor fear-avoidance questionnaire and its association with pain intensity, catastrophizing, and disability. Journal of Burn Care & Research 2023:irad025.

[25] Cheung PWH, Wong CKH, Cheung JPY. Psychometric validation of the cross-culturally adapted traditional Chinese version of the Back Beliefs Questionnaire (BBQ) and Fear-Avoidance Beliefs Questionnaire (FABQ). European Spine Journal 2018;27:1724-1733.

[26] Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. Journal of Clinical Epidemiology 2018;95:73-93.

[27] Cho S, Lee S-M, McCracken LM, Moon D-E, Heiby EM. Psychometric properties of a Korean version of the Pain Anxiety Symptoms Scale-20 in chronic pain patients. International journal of behavioral medicine 2010;17:108-117.

[28] Cipher DJ, Fernandez E. Expectancy variables predicting tolerance and avoidance of pain in chronic pain patients. Behaviour Research and Therapy 1997;35(5):437-444.

[29] Claes N, Crombez G, Vlaeyen JW. Pain-avoidance versus reward-seeking: an experimental investigation. Pain 2015;156(8):1449-1457.

[30] Claes N, Karos K, Meulders A, Crombez G, Vlaeyen JWS. Competing Goals Attenuate Avoidance Behavior in the Context of Pain. Journal of Pain 2014;15(11):1120-1129.

[31] Claes N, Vlaeyen JWS, Crombez G. Pain in context: Cues predicting a reward decrease fear of movement related pain and avoidance behavior. Behaviour Research and Therapy 2016;84:35-44.

[32] Clark M. Kinesiophobia and chronic pain: psychometric characteristics and factor analysis of the Tampa Scale, Proceedings of the Paper presented at: The 15th Annual Scientific Meeting of the American Pain Society; 1996; Washington, DC, Vol. 77, 1996.

[33] Cleland JA, Fritz JM, Brennan GP. Predictive validity of initial fear avoidance beliefs in patients with low back pain receiving physical therapy: is the FABQ a useful screening tool for identifying patients at risk for a poor recovery? European Spine Journal 2008;17:70-79.

[34] Cleland JA, Fritz JM, Childs JD. Psychometric properties of the Fear-Avoidance Beliefs Questionnaire and Tampa Scale of Kinesiophobia in patients with neck pain. American journal of physical medicine & rehabilitation 2008;87(2):109-117.

[35] Coons MJ, Hadjistavropoulos HD, Asmundson GJ. Factor structure and psychometric properties of the Pain Anxiety Symptoms Scale-20 in a community physiotherapy clinic sample. European journal of pain 2004;8(6):511-516.

[36] Cordeiro N, Pezarat-Correia P, Gil J, Cabri J. Portuguese language version of the Tampa Scale for Kinesiophobia [13 items]. Journal of Musculoskeletal Pain 2013;21(1):58-63.

[37] Cornelius T, Schenk P, Dixon D, Crombez G, Johnston M. Double, Double, Measurement Trouble, Sorry to Burst your Construct Bubble. European Health Psychologist 2024;23(2).

[38] Crombez G, Vlaeyen JW, Heuts PH, Lysens R. Pain-related fear is more disabling than pain itself: evidence on the role of pain-related fear in chronic back pain disability. Pain 1999;80(1-2):329-339.

[39] Crul B, Crombez G, Vlaeyen JW. Meten en beoordelen van pijnproblemen bij patiënten: hoe en waarom? Boekblok Pijn info 2009:144-148.

[40] Cuesta-Vargas AI, Neblett R, Gatchel RJ, Roldán-Jiménez C. Cross-cultural adaptation and validity of the Spanish fear-avoidance components scale and clinical implications in primary care. BMC family practice 2020;21:1-9.

[41] Dannecker EA, George SZ. A Comparison of Laboratory Measures of Escape and Avoidance Behavior. Journal of Pain 2009;10(1):53-59.

[42] Dayalan H, Aseer P. Translation, Cross-Cultural Adaptation, Reliability, and Validity of the Tamil Version of Fear-Avoidance Beliefs Questionnaire in Chronic Low Back Pain. Internet Journal of Allied Health Sciences and Practice 2022;20(4):11.

[43] De Baets L, Sergooris A, Neblett R, Matheve T, Mingels S, Van Goethem A, Huybrechts X, Corten K, Gerits D, Vandevoort D. The development and measurement properties of the Dutch version of the fear-avoidance components scale (FACS-D) in persons with chronic musculoskeletal pain. Scandinavian Journal of Pain 2023;23(2):298-307.

[44] De Souza FS, da Silva Marinho C, Siqueira FB, Maher CG, Costa LOP. Psychometric testing confirms that the Brazilian-Portuguese adaptations, the original versions of the Fear-Avoidance Beliefs Questionnaire, and the Tampa Scale of Kinesiophobia have similar measurement properties. Spine 2008;33(9):1028-1033.

[45] Dedering Å, Börjesson T. Assessing Fear-avoidance Beliefs in Patients with Cervical Radiculopathy. Physiotherapy Research International 2013;18(4):193-202.

[46] Duport A, Bédard S, Raynauld C, Bordeleau M, Neblett R, Balg F, Devanne H, Léonard G. Cross-cultural translation and psychometric validation of the French version of the Fear-Avoidance Components Scale (FACS). Plos one 2023;18(10):e0288899.

[47] Dupuis F, Cherif A, Batcho C, Massé-Alarie H, Roy J-S. The Tampa scale of kinesiophobia: A systematic review of its psychometric properties in people with musculoskeletal pain. The Clinical journal of pain 2023;39(5):236-247.

[48] Eiger B, Errebo M, Straszek CL, Vaegter HB. Less is more: reliability and measurement error for three versions of the Tampa Scale of Kinesiophobia (TSK-11, TSK-13, and TSK-17) in patients with high-impact chronic pain. Scandinavian journal of pain 2023;23(1):217-224.

[49] Elsman EB, Butcher NJ, Mokkink LB, Terwee CB, Tricco A, Gagnier JJ, Aiyegbusi OL, Barnett C, Smith M, Moher D. Study protocol for developing, piloting and disseminating the PRISMA-COSMIN guideline: a new reporting guideline for systematic reviews of outcome measurement instruments. Systematic Reviews 2022;11(1):121.

[50] Esteve R, López-Martínez AE, Peters ML, Serrano-Ibáñez ER, Ruiz-Párraga GT, Ramírez-Maestre C. Optimism, Positive and Negative Affect, and Goal Adjustment Strategies: Their Relationship to Activity Patterns in Patients with Chronic Musculoskeletal Pain. Pain Res Manag 2018;2018:6291719.

[51] Esteve R, Ramirez-Maestre C, Peters ML, Serrano-Ibanez ER, Ruiz-Parraga GT, Lopez-Martinez AE. Development and initial validation of the activity patterns scale in patients with chronic pain. The Journal of Pain 2016;17(4):451-461.

[52] Franchignoni F, Giordano A, Rocca B, Ferriero G, Monticone M. A further Rasch analysis of the Fear-Avoidance Beliefs Questionnaire in adults with chronic low back pain suggests the revision of its rating scale. European Journal of Physical and Rehabilitation Medicine 2020;57(1):110-119.

[53] French DJ, France CR, Vigneau F, French JA, Evans RT. Fear of movement/(re) injury in chronic pain: a psychometric assessment of the original English version of the Tampa scale for kinesiophobia (TSK). Pain 2007;127(1-2):42-51.

[54] Fritz JM, George SZ. Identifying psychosocial variables in patients with acute work-related low back pain: the importance of fear-avoidance beliefs. Phys Ther 2002;82(10):973-983.

[55] Garcia-Campayo J, Pascual A, Alda M, Ramirez MTG. Coping with fibromialgia: usefulness of the Chronic Pain Coping Inventory-42. Pain 2007;132:S68-S76.

[56] Geisser ME, Haig AJ, Theisen ME. Activity avoidance and function in persons with chronic back pain. Journal of occupational rehabilitation 2000;10:215-227.

[57] Gendreau M, Hufford MR, Stone AA. Measuring clinical pain in chronic widespread pain: selected methodological issues. Best Practice & Research Clinical Rheumatology 2003;17(4):575-592.

[58] George SZ, Fritz JM, McNeil DW. Fear-avoidance beliefs as measured by the Fear-Avoidance Beliefs Questionnaire: Change in Fear-Avoidance Beliefs Questionnaire is predictive of change in self-report of disability and pain intensity for patients with acute low back pain. Clinical Journal of Pain 2006;22(2):197-203.

[59] George SZ, Valencia C, Beneciuk JM. A psychometric investigation of fear-avoidance model measures in patients with chronic low back pain. journal of orthopaedic & sports physical therapy 2010;40(4):197-205.

[60] Georgoudis G, Papathanasiou G, Spiropoulos P, Katsoulakis K. Cognitive assessment of musculoskeletal pain with a newly validated Greek version of the Fear-Avoidance Beliefs Questionnaire (FABQ). European journal of pain 2007;11(3):341-351.

[61] Georgoudis G, Raptis K, Koutserimpas C. Cognitive Assessment of Musculoskeletal Pain: Validity and Reliability of the Greek Version of the Tampa Scale of Kinesiophobia in Patients Suffering from Chronic Low Back Pain. Maedica 2022;17(4):826.

[62] González Aroca J, Díaz ÁP, Navarrete C, Albarnez L. Fear-Avoidance Beliefs Are Associated with Pain Intensity and Shoulder Disability in Adults with Chronic Shoulder Pain: A Cross-Sectional Study. Journal of Clinical Medicine 2023;12(10):3376.

[63] Goubert L, Crombez G, Van Damme S, Vlaeyen JW, Bijttebier P, Roelofs J. Confirmatory factor analysis of the Tampa Scale for Kinesiophobia: invariant two-factor model across low back pain patients and fibromyalgia patients. The Clinical journal of pain 2004;20(2):103-110.

[64] Grotle M, Brox JI, Vøllestad NK. Reliability, validity and responsiveness of the fear-avoidance beliefs questionnaire: methodological aspects of the Norwegian version. Journal of rehabilitation medicine 2006;38(6):346-353.

[65] Gómez-Pérez L, López-Martínez AE, Ruiz-Párraga GT. Psychometric properties of the Spanish version of the Tampa Scale for Kinesiophobia (TSK). The journal of Pain 2011;12(4):425-435.

[66] Hadjistavropoulos HD, MacLeod FK, Asmundson GJ. Validation of the chronic pain coping inventory. Pain 1999;80(3):471-481.

[67] Hapidou EG, O'Brien MA, Pierrynowski MR, de Las Heras E, Patel M, Patla T. Fear and avoidance of movement in people with chronic pain: psychometric properties of the 11-item Tampa Scale for Kinesiophobia (TSK-11). Physiotherapy Canada 2012;64(3):235-241.

[68] Hasenbring MI, Hallner D, Rusu AC. Fear-avoidance-and endurance-related responses to pain: development and validation of the Avoidance-Endurance Questionnaire (AEQ). European Journal of Pain 2009;13(6):620-628.

[69] Haugen AJ, Grøvle L, Keller A, Grotle M. Cross-cultural adaptation and validation of the Norwegian version of the Tampa scale for kinesiophobia. Spine 2008;33(17):E595-E601.

[70] He S, Wang J, Ji P. Validation of the Tampa Scale for Kinesiophobia for Temporomandibular Disorders (TSK-TMD) in patients with painful TMD. The journal of headache and pain 2016;17:1-5.

[71] Heuts PH, Vlaeyen JW, Roelofs J, de Bie RA, Aretz K, van Weel C, van Schayck OC. Pain-related fear and daily functioning in patients with osteoarthritis. Pain 2004;110(1-2):228-235.

[72] Holzapfel S, Riecke J, Rief W, Schneider J, Glombiewski JA. Development and validation of the Behavioral Avoidance Test-Back pain (BAT-back) for patients with chronic low back pain. Clinical Journal of Pain 2016;32(11):940-947.

[73] Ibrahim AA, Akindele MO, Kaka B, Bello B. Translation, cross-cultural adaptation, and psychometric properties of the Hausa version of the Fear-Avoidance Beliefs Questionnaire in patients with low back pain. Scandinavian Journal of Pain 2019;19(1):83-92.

[74] Igwesi-Chidobe CN, Amarachukwu C, Sorinola IO, Godfrey EL. Translation, cultural adaptation and psychometric testing of Igbo fear avoidance beliefs questionnaire in mixed rural and urban Nigerian populations with chronic low back pain. PLoS One 2019;14(5):e0216482.

[75] Inrig T, Amey B, Borthwick C, Beaton D. Validity and reliability of the Fear-Avoidance Beliefs Questionnaire (FABQ) in workers with upper extremity injuries. Journal of occupational rehabilitation 2012;22:59-70.

[76] Jakobsson M, Gutke A, Mokkink LB, Smeets R, Lundberg M. Level of evidence for reliability, validity, and responsiveness of physical capacity tasks designed to assess functioning in patients

with low back pain: a systematic review using the COSMIN standards. Physical therapy 2019;99(4):457-477.

[77] Jaspers ME, van Haasterecht L, van Zuijlen PP, Mokkink LB. A systematic review on the quality of measurement techniques for the assessment of burn wound depth or healing potential. Burns 2019;45(2):261-281.

[78] Jensen MP, Turner JA, Romano JM, Strom SE. The chronic pain coping inventory: development and preliminary validation. Pain 1995;60(2):203-216.

[79] Jensen MP, Ward LC, Thorn BE, Ehde DM, Day MA. Measuring the cognitions, emotions, and motivation associated with avoidance behaviors in the context of pain: preliminary development of the negative responsivity to pain scales. The Clinical Journal of Pain 2017;33(4):325-334.

[80] Kaka B, Ogwumike OO, Idowu OA, Odole AC, Saidu AM, Fawole HO, Ibrahim M. Translation of the fear avoidance beliefs questionnaire into Hausa language. Global journal of health science 2015;7(3):116.

[81] Kamonseki DH, Haik MN, Ribeiro LP, Almeida RFd, Almeida LAd, Fonseca CL, Camargo PR. Measurement properties of the Brazilian versions of Fear-Avoidance Beliefs Questionnaire and Tampa Scale of Kinesiophobia in individuals with shoulder pain. PloS one 2021;16(12):e0260452.

[82] Karsdorp PA, Geenen R, Kroese FM, Vlaeyen JWS. Turning Pain Into Cues for Goal-Directed Behavior: Implementation Intentions Reduce Escape-Avoidance Behavior, on a Painful Task. Journal of Pain 2016;17(4):499-507.

[83] Karsdorp PA, Nijst SE, Goossens ME, Vlaeyen JW. The role of current mood and stop rules on physical task performance: an experimental investigation in patients with work-related upper extremity pain. Eur J Pain 2010;14(4):434-440.

[84] Karsdorp PA, Vlaeyen JW. Goals matter: both achievement and pain-avoidance goals are associated with pain severity and disability in patients with low back and upper extremity pain. Pain 2011;152(6):1382-1390.

[85] Kemani M, Zetterqvist V, Kanstrup M, Holmström L, Wicksell R. A validation of the pain interference index in adults with long-standing pain. Acta Anaesthesiologica Scandinavica 2016;60(2):250-258.

[86] Kikuchi N, Matsudaira K, Sawada T, Oka H. Psychometric properties of the Japanese version of the Tampa Scale for Kinesiophobia (TSK-J) in patients with whiplash neck injury pain and/or low back pain. Journal of orthopaedic science 2015;20:985-992.

[87] Knezevic A, Neblett R, Gatchel RJ, Jeremic-Knezevic M, Bugarski-Ignjatovic V, Tomasevic-Todorovic S, Boskovic K, Cuesta-Vargas AI. Psychometric validation of the Serbian version of the Fear Avoidance Component Scale (FACS). PLoS One 2018;13(9):e0204311.

[88] Ko Y-M, Park W-B, Lim J-Y. Cross-cultural adaptation and clinimetric property of Korean version of the Chronic Pain Coping Inventory-42 in patients with chronic low back pain. Spine 2010;35(6):666-671.

[89] Koho P, Aho S, Kautiainen H, Pohjolainen T, Hurri H. Test–retest reliability and comparability of paper and computer questionnaires for the Finnish version of the Tampa Scale of Kinesiophobia. Physiotherapy 2014;100(4):356-362.

[90] Korkmaz N, Akinci A, Yörükan S, Sürücü HS, Saraçbaşi O, Ozçakar L. Validation and reliability of the Turkish version of the fear avoidance beliefs questionnaire in patients with low back pain. European journal of physical and rehabilitation medicine 2009;45(4):527-535.

[91] Kovacs FM, Muriel A, Medina JM, Abraira V, Sánchez MDC, Jaúregui JO, Network SBPR. Psychometric characteristics of the Spanish version of the FAB questionnaire. Spine 2006;31(1):104-110.

[92] Kraaimaat F, Bakker A, Evers A. Pijncoping-strategieën bij chronische pijnpatiënten: De ontwikkeling van de Pijn-Coping-Inventarisatielijst (PCI). 1997.

[93] Kraaimaat FW, Evers AW. Pain-coping strategies in chronic pain patients: psychometric characteristics of the pain-coping inventory (PCI). International journal of behavioral medicine 2003;10:343-363.

[94] Kreddig N, Rusu AC, Burkhardt K, Hasenbring MI. The German PASS-20 in patients with low back pain: new aspects of convergent, divergent, and criterion-related validity. International journal of behavioral medicine 2015;22:197-205.

[95] Krypotos A-M, Vervliet B, Engelhard IM. The validity of human avoidance paradigms. Behaviour Research and Therapy 2018;111:99-105.

[96] Krypotos AM, Crombez G, Vlaeyen JWS. The dynamics of pain avoidance: the exploration-exploitation dilemma. Pain 2024;165(10):2145-2149.

[97] Krypotos AM, Sjouwerman R, Teppers M, Vlaeyen JWS. Pavlovian-to-instrumental transfer in individuals with chronic pain. Behav Res Ther 2024;176:104491.

[98] Kugler K, Wijn J, Geilen M, De Jong J, Vlaeyen J. The photograph series of daily activities (PHODA). CD-rom version 1999;1.

[99] Kumar A, Pithadia K, Kumar D, Sannasi R. Cross-cultural adaptation, reliability, and validity of Kannada version of Fear Avoidance Belief Questionnaire in chronic low back pain. Revista Pesquisa em Fisioterapia 2020;10(4):610-618.

[100] Küçükakkaş O, Karaman ÇA. Cross-cultural adaptation and validation of the behavioral avoidance test-Back pain (BAT-Back) to the Turkish language. Journal of Orthopaedic Science 2020;25(2):219-223.

[101] La Touche R, Pardo-Montero J, Cuenca-Martínez F, Visscher CM, Paris-Alemany A, López-de-Uralde-Villanueva I. Cross-cultural adaptation and psychometric properties of the Spanish version of the Tampa Scale for Kinesiophobia for temporomandibular disorders. Journal of Clinical Medicine 2020;9(9):2831.

[102] Lamé IE, Peters ML, Kessels AG, Van Kleef M, Patijn J. Test—retest stability of the pain catastrophizing scale and the tampa scale for kinesiophobia in chronic pain over a longer period of time. Journal of health psychology 2008;13(6):820-826.

[103] Langlois J, Vincent-Toskin S, Duchesne P, de Vilhena BS, Shashoua D, Calva V, de Oliveira A, Nedelec B. Fear avoidance beliefs and behaviors of burn survivors: a mixed methods approach. Burns 2021;47(1):175-189.

[104] Larsson C, Hansson EE, Sundquist K, Jakobsson U. Psychometric properties of the Tampa Scale of Kinesiophobia (TSK-11) among older people with chronic pain. Physiotherapy theory and practice 2014;30(6):421-428.

[105] Laufer Y, Elheiga-Na'amne BA, Rozen N. Translation and validation of the Arab version of the fear avoidance beliefs questionnaire. Journal of back and musculoskeletal rehabilitation 2012;25(3):201-208.

[106] Lee K-C, Chiu TT, Lam T-H. Psychometric properties of the Fear-Avoidance Beliefs Questionnaire in patients with neck pain. Clinical rehabilitation 2006;20(10):909-920.

[107] Leeuw M, Goossens ME, Linton SJ, Crombez G, Boersma K, Vlaeyen JW. The fear-avoidance model of musculoskeletal pain: current state of scientific evidence. Journal of Behavioral Medicine 2007;30(1):77-94.

[108] Leeuw M, Goossens MEJB, van Breukelen GJP, Boersma K, Vlaeyen JWS. Measuring Perceived Harmfulness of Physical Activities in Patients With Chronic Low Back Pain: The Photograph Series of Daily Activities-Short Electronic Version. Journal of Pain 2007;8(11):840-849.

[109] Lorente S, Viladrich C, Vives J, Losilla J-M. Tools to assess the measurement properties of quality of life instruments: a meta-review. BMJ open 2020;10(8):e036038.

[110] Lundberg MK, Styf J, Carlsson SG. A psychometric evaluation of the Tampa Scale for Kinesiophobia—from a physiotherapeutic perspective. Physiotherapy theory and practice 2004;20(2):121-133.

[111] Matsudaira K, Kikuchi N, Murakami A, Isomura T. Psychometric properties of the Japanese version of the Fear-Avoidance Beliefs Questionnaire (FABQ). Journal of Orthopaedic Science 2014;19:26-32.

[112] Mbada C, Idowu O, Awosunle G, Adeniyi A, Oke K, Johnson O, Odole A. Translation, cultural adaptation, and psychometric testing of the Yoruba version of Fear-Avoidance Beliefs Questionnaire in patients with low-back pain. Disability and rehabilitation 2021;43(6):846-852.

[113] McCracken LM, Dhingra L. A short version of the Pain Anxiety Symptoms Scale (PASS-20): Preliminary development and validity. Pain research and management 2002;7(1):45-50.

[114] McCracken LM, Samuel VM. The role of avoidance, pacing, and other activity patterns in chronic pain. Pain 2007;130(1-2):119-125.

[115] McCracken LM, Vowles KE, Eccleston C. Chronic Pain Acceptance Questionnaire–Revised (CPAQ-R). Acceptance and Commitment Therapy Measures Package 2006:146.

[116] McCracken LM, Zayfert C, Gross RT. The Pain Anxiety Symptoms Scale: development and validation of a scale to measure fear of pain. Pain 1992;50(1):67-73.

[117] McNeish D. Thanks coefficient alpha, we'll take it from here. Psychological methods 2018;23(3):412.

[118] Meulders A. From fear of movement-related pain and avoidance to chronic pain disability: a state-of-the-art review. Current Opinion in Behavioral Sciences 2019;26:130-136.

[119] Meulders A, Franssen M, Fonteyne R, Vlaeyen JWS. Acquisition and extinction of operant pain-related avoidance behavior using a 3 degrees-of-freedom robotic arm. Pain 2016;157(5):1094-1104.

[120] Meulders A, Vansteenwegen D, Vlaeyen JW. The acquisition of fear of movement-related pain and associative learning: a novel pain-relevant human fear conditioning paradigm. Pain 2011;152(11):2460-2469.

[121] Miller R, Kori S, Todd D. The tampa scale. Unpublished report. Tampa, FL 1991.

[122] Mintken PE, Cleland JA, Whitman JM, George SZ. Psychometric properties of the Fear-Avoidance Beliefs Questionnaire and Tampa Scale of Kinesiophobia in patients with shoulder pain. Archives of physical medicine and rehabilitation 2010;91(7):1128-1136.

[123] Misterska E, Jankowski R, Głowacki M. Psychometric properties of the Polish language version of the chronic pain coping inventory-42 for patients treated surgically due to herniated lumbar discs and spondylotic changes. Medical Science Monitor: International Medical Journal of Experimental and Clinical Research 2014;20:789.

[124] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine 2009;151(4):264-269.

[125] Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter LM, De Vet H, Terwee CB, Mokkink L. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User manual 2018;78(1).

[126] Mokkink LB, Prinsen CA, Bouter LM, de Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Brazilian journal of physical therapy 2016;20(2):105-113.

[127] Monticone M, Ambrosini E, Rocca B, Foti C, Ferrante S. Responsiveness and minimal clinically important changes for the Tampa Scale of Kinesiophobia after lumbar fusion during cognitive behavioral rehabilitation. European journal of physical and rehabilitation medicine 2016;53(3):351-358.

[128] Monticone M, Ambrosini E, Rocca B, Foti C, Ferrante S. Responsiveness of the Tampa Scale of Kinesiophobia in Italian subjects with chronic low back pain undergoing motor and cognitive rehabilitation. European spine journal 2016;25:2882-2888.

[129] Monticone M, Baiardi P, Bonetti F, Ferrari S, Foti C, Pillastrini P, Rocca B, Vanti C, Zanoli G. The Italian version of the Fear-Avoidance Beliefs Questionnaire (FABQ-I): cross-cultural adaptation, factor analysis, reliability, validity, and sensitivity to change. Spine 2012;37(6):E374-E380.

[130] Monticone M, Ferrante S, Giorgi I, Galandra C, Rocca B, Foti C. Development of the Italian version of the 42-item Chronic Pain Coping Inventory, CPCI-I: Cross-cultural adaptation, factor analysis, reliability and validity. Quality of Life Research 2013;22:1459-1465.

[131] Monticone M, Frigau L, Vernon H, Rocca B, Giordano A, Vullo SS, Mola F, Franchignoni F. Reliability, responsiveness and minimal clinically important difference of the two Fear Avoidance and Beliefs Questionnaire scales in Italian subjects with chronic low back pain undergoing multidisciplinary rehabilitation. European journal of physical and rehabilitation medicine 2020.

[132] Monticone M, Giorgi I, Baiardi P, Barbieri M, Rocca B, Bonezzi C. Development of the Italian version of the Tampa Scale of Kinesiophobia (TSK-I): cross-cultural adaptation, factor analysis, reliability, and validity. Spine 2010;35(12):1241-1246.

[133] Nagasawa Y, Shibata A, Fukamachi H, Ishii K, Wicksell RK, Oka K. The Psychological Inflexibility in Pain Scale (PIPS): validity and reliability of the Japanese version for chronic low back pain and knee pain. Journal of pain research 2021:325-332.

[134] Namli M, KESİKTAŞ FN, Paker N, ŞİRİN B, ALAYOĞLU FO, Ersoy S. THE RELATIONSHIP BETWEEN FEAR AVOIDANCE BELIEFS AND PHYSICAL DISABILITY IN PATIENTS WITH SYMPTOMATIC KNEE OSTEOARTHRITIS. Turkish Journal of Geriatrics/Türk Geriatri Dergisi 2022;25(2).

[135] Neblett R, Mayer TG, Hartzell MM, Williams MJ, Gatchel RJ. The Fear-avoidance components scale (FACS): development and psychometric evaluation of a new measure of pain-related fear avoidance. Pain Practice 2016;16(4):435-450.

[136] Neblett R, Mayer TG, Williams MJ, Asih S, Cuesta-Vargas AI, Hartzell MM, Gatchel RJ. The fear-avoidance components scale (FACS): responsiveness to functional restoration treatment in a chronic musculoskeletal pain disorder (CMPD) population. The Clinical Journal of Pain 2017;33(12):1088-1099.

[137] Nees F, Ruttorf M, Fuchs X, Rance M. Approach and Avoidance Behavior in Chronic Back Pain: The Role of Pavlovian-Instrumental Transfer. Psychophysiology 2015;52:S75-S75.

[138] Ozuberk B, Aslan U, Ekici G. The validity and reliability of Turkish adaptation of the Fear-Avoidance Beliefs Questionnaire in patients with neck pain. Northern Clinics of Istanbul 2023;10(3).

[139] Pagels L, Lüdtke K, Schäfer A. Validation of the German version of the Fear-Avoidance Beliefs Questionnaire (FABQ-D) for shoulder disorders. Schmerz (Berlin, Germany) 2023.

[140] Panhale VP, Gurav RS, Shah P, Nayak N. Psychometric properties of the Hindi version of the fear-Avoidance beliefs questionnaire in patients with chronic non-specific low back pain. Spine 2019;44(15):E908-E913.

[141] Panhale VP, Gurav RS, Suradkar K. Translation and validation of Marathi version of Fear-Avoidance and Belief Questionnaire in patients with chronic low back pain. Indian Journal of Pain 2018;32(3):173-178.

[142] Pei L, Xia J, Yan J. Cross-cultural adaptation, reliability and validity of the Chinese version of the Fear Avoidance Beliefs Questionnaire. Journal of International Medical Research 2010;38(6):1985-1996.

[143] Peñacoba C, López-Gómez I, Pastor-Mira MA, López-Roig S, Ecija C. Contextualizing goal preferences in fear-avoidance models. Looking at fatigue as a disabling symptom in fibromyalgia patients. PLoS One 2021;16(7):e0254200.

[144] Pfingsten M. Angstvermeidungs-Überzeugungen bei Rückenschmerzen: Gütekriterien und prognostische Relevanz des FABQ. Der Schmerz 2004;18:17-27.

[145] Pfingsten M, Kröner-Herwig B, Leibing E, Kronshage U. Validation of the German version of the fear-avoidance beliefs questionnaire (FABQ). European journal of pain 2000;4(3):259-266.

[146] Pfingsten M, Leibing E, Franz C, Bansemer D, Busch O, Hildebrandt J. Fear-avoidance-beliefs in patients with backpain: Deutsche Version des „fear-avoidance-beliefs questionnaire"(FABQ-D). Der Schmerz 1997;11:387-395.

[147] Pfingsten M, Leibing E, Harter W, Kröner-Herwig B, Hempel D, Kronshage U, Hildebrandt J. Fear-avoidance behavior and anticipation of pain in patients with chronic low back pain: A randomized controlled study. Pain Medicine 2001;2(4):259-266.

[148] Philips H. Avoidance behaviour and its role in sustaining chronic pain. Behaviour research and therapy 1987;25(4):273-279.

[149] Pincus T, Smeets RJ, Simmonds MJ, Sullivan MJ. The fear avoidance model disentangled: improving the clinical utility of the fear avoidance model. The Clinical journal of pain 2010;26(9):739-746.

[150] Pittig A, Wong AH, Glück VM, Boschet JM. Avoidance and its bi-directional relationship with conditioned fear: Mechanisms, moderators, and clinical implications. Behaviour Research and Therapy 2020:103550.

[151] Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, De Vet HC, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research 2018;27(5):1147-1157.

[152] Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, Williamson PR, Terwee CB. How to select outcome measurement instruments for outcomes included in a “Core Outcome Set”–a practical guideline. Trials 2016;17(1):449.

[153] Prkachin KM, Schultz IZ, Hughes E. Pain behavior and the development of pain-related disability: the importance of guarding. Clin J Pain 2007;23(3):270-277.

[154] Pruneti C, Sgromo D, Bicchieri L, Fontana F, Franceschini M, Ferraro F, Cosentino C. Contribution to the validation of Italian version of Fear-Avoidance Beliefs Questionnaire. Acta Bio Medica Atenei Parmensis 2014;85(1):8-17.

[155] Pulles AN, Köke AJ, Strackke RP, Smeets RJ. The responsiveness and interpretability of psychosocial patient-reported outcome measures in chronic musculoskeletal pain rehabilitation. European Journal of Pain 2020;24(1):134-144.

[156] Quint S, Luckmann J, Wolf U, Basler HD. AMIKA: Psychometric evaluation of a photo-based scale for the assessment of fear avoidance beliefs in elderly individuals. Schmerz 2007;21(5):453-461.

[157] Quint S, Raich M, Luckmann J. Evaluation einer zweidimensionalen Skala zur Erfassung von Angst-Vermeidungs-Überzeugungen an älteren Patienten mit chronischem Rückenschmerz. Der Schmerz 2011;3(25):315-321.

[158] Raich MMJ. Erfassung von Fear-Avoidance Beliefs bei älteren Patienten mit chronischem Rückenschmerz durch zwei Kurzfragebögen: AMIKA-K & KVS-D-65+. Fachbereich Humanmedizin Philipps-Universität, Marburg, S 2009;147.

[159] Reneman MF, Kleen M, Trompetter HR, Preuper HRS, Köke A, Van Baalen B, Schreurs KM. Measuring avoidance of pain: validation of the Acceptance and Action Questionnaire II-pain version. International Journal of Rehabilitation Research 2014;37(2):125-129.

[160] Riley J, Ahern D, Follick M. Chronic pain and functional impairment: assessing beliefs about their relationship. Archives of physical medicine and rehabilitation 1988;69(8):579-582.

[161] Riley SP, Tafuto V, Cote M, Brismée J-M, Wright A, Cook C. Reliability and relationship of the fear-avoidance beliefs questionnaire with the shoulder pain and disability index and numeric

pain rating scale in patients with shoulder pain. Physiotherapy Theory and Practice 2019;35(5):464-470.

[162] Rodero B, Pereira JP, Pérez-Yus MC, Casanueva B, Serrano-Blanco A, Rodrigues da Cunha Ribeiro MJ, Luciano JV, Garcia-Campayo J. Validation of a Spanish version of the psychological inflexibility in pain scale (PIPS) and an evaluation of its relation with acceptance of pain and mindfulness in sample of persons with fibromyalgia. Health and Quality of Life Outcomes 2013;11:1-10.

[163] Roelofs J, Goubert L, Peters ML, Vlaeyen JW, Crombez G. The Tampa Scale for Kinesiophobia: further examination of psychometric properties in patients with chronic low back pain and fibromyalgia. European Journal of Pain 2004;8(5):495-502.

[164] Roelofs J, Sluiter JK, Frings-Dresen MH, Goossens M, Thibault P, Boersma K, Vlaeyen JW. Fear of movement and (re) injury in chronic musculoskeletal pain: Evidence for an invariant two-factor model of the Tampa Scale for Kinesiophobia across pain diagnoses and Dutch, Swedish, and Canadian samples. Pain 2007;131(1-2):181-190.

[165] Roelofs J, van Breukelen G, Sluiter J, Frings-Dresen MHW, Goossens M, Thibault P, Boersma K, Vlaeyen JWS. Norming of the Tampa Scale for Kinesiophobia across pain diagnoses and various countries. Pain 2011;152(5):1090-1095.

[166] Romano JM, Jensen MP, Turner JA. The chronic pain coping inventory-42: reliability and validity. Pain 2003;104(1-2):65-73.

[167] Rosenkoetter U, Tate RL. Assessing features of psychometric assessment instruments: a comparison of the COSMIN checklist with other critical appraisal tools. Brain Impairment 2018;19(1):103-118.

[168] Rosenstiel AK, Keefe FJ. The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. Pain 1983;17(1):33-44.

[169] Rostami M, Noorian N, Mansournia MA, Sharafi E, Babaki AES, Kordi R. Validation of the Persian version of the fear avoidance belief questionnaire in patients with low back pain. Journal of back and musculoskeletal rehabilitation 2014;27(2):213-221.

[170] Ruiz-Párraga GT, López-Martínez AE, Rusu AC, Hasenbring MI. Spanish version of the avoidance-endurance questionnaire: factor structure and psychometric properties. The Spanish Journal of Psychology 2015;18:E88.

[171] Rusu AC, Kreddig N, Hallner D, Hülsebusch J, Hasenbring MI. Fear of movement/(Re) injury in low back pain: confirmatory validation of a German version of the Tampa Scale for Kinesiophobia. BMC musculoskeletal disorders 2014;15:1-9.

[172] Saadat M, Salamat S, Mostafaee N, Soleimani F, Rouintan Z, Amin M. To evaluate responsiveness and minimal important change (MIC) for the Persian versions of FABQ, TSK, and PCS. European Spine Journal 2023;32(9):3023-3029.

[173] Salama HM, Reda N, El Shahaly M, Nour-Eldein H. Predictors of fear-avoidance belief, pain, and disability index in patients with chronic low back pain attending rheumatology outpatient clinics. Journal of Public Health 2020:1-6.

[174] Santo Salvador EME, Franco KFM, Miyamoto GC, dos Santos Franco YR, Cabral CMN. Analysis of the measurement properties of the Brazilian-Portuguese version of the Tampa Scale for Kinesiophobia-11 in patients with fibromyalgia. Brazilian Journal of Physical Therapy 2021;25(2):168-174.

[175] Satpute KH, Ranade PS, Hall TM. Development of the Marathi version of the Tampa scale of kinesiophobia 11: Cross-cultural adaptation, validity, and test–retest reliability in patients with low back pain. Indian Spine Journal 2019;2(2):146-151.

[176] Sege CT, Bradley MM, Lang PJ. Avoidance and escape: Defensive reactivity and trait anxiety. Behaviour Research and Therapy 2018;104:62-68.

[177] Siqueira FB, Teixeira-Salmela LF, Magalhães LdC. Analysis of the psychometric properties of the Brazilian version the tampa scale for kinesiophobia. Acta Ortopédica Brasileira 2007;15:19-24.

[178] Sorbi MJ, Peters ML, Kruise DA, Maas CJ, Kerssens JJ, Verhaak PF, Bensing JM. Electronic momentary assessment in chronic pain I: psychological pain responses as predictors of pain intensity. The Clinical journal of pain 2006;22(1):55-66.

[179] Souza LAF, Cruz DdALMd, Pereira LV. Cross-cultural adaptation of Chronic Pain Coping Inventory-Brazilian version. BrJP 2018;1:94-102.

[180] Souza LAF, Pereira LV, de Moura LA, Díaz L-JR, da Cruz DdALM, Aparecido Da Silva J. Structural validity of the Chronic Pain Coping Inventory—Brazilian version. Plos one 2021;16(2):e0246294.

[181] Speyer R, Cordier R, Kertscher B, Heijnen BJ. Psychometric properties of questionnaires on functional health status in oropharyngeal dysphagia: a systematic literature review. BioMed research international 2014;2014(1):458678.

[182] Staerkle R, Mannion AF, Elfering A, Junge A, Semmer NK, Jacobshagen N, Grob D, Dvorak J, Boos N. Longitudinal validation of the fear-avoidance beliefs questionnaire (FABQ) in a Swiss-German sample of low back pain patients. European Spine Journal 2004;13:332-340.

[183] Sørensen L, van Tulder M, Johannsen HV, Ovesen J, Oestergaard LG. Responsiveness and minimal important change of the Oxford Shoulder Score, EQ-5D, and the Fear-Avoidance Belief Questionnaire Physical Activity subscale in patients undergoing arthroscopic subacromial decompression. JSES international 2021;5(5):869-874.

[184] Tan G, Nguyen Q, Anderson KO, Jensen M, Thornby J. Further validation of the chronic pain coping inventory. The Journal of Pain 2005;6(1):29-40.

[185] Terho H, Haapea M, Paananen M, Korniloff K, Häkkinen A, Karppinen J. Translation and validation of the Finnish version of the Fear-Avoidance Beliefs Questionnaire (FABQ). Scandinavian journal of pain 2016;10(1):113-118.

[186] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of clinical epidemiology 2007;60(1):34-42.

[187] Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of life research 2012;21(4):651-657.

[188] Tkachuk GA, Harris CA. Psychometric properties of the Tampa Scale for Kinesiophobia-11 (TSK-11). The Journal of Pain 2012;13(10):970-977.

[189] Trolle N, Christiansen DH. Measurement properties of the Fear-Avoidance Belief Questionnaire for physical activity in patients with shoulder impingement syndrome. Patient Related Outcome Measures 2019:83-87.

[190] Trompetter HR, Bohlmeijer ET, Van Baalen B, Kleen M, Köke A, Reneman M, Schreurs KM. The psychological inflexibility in pain scale (pips). European journal of psychological assessment 2014.

[191] Truchon M, Côté D, Irachabal S. The Chronic Pain Coping Inventory: confirmatory factor analysis of the French version. BMC Musculoskeletal Disorders 2006;7:1-9.

[192] Turan K, Sarı Z, Özden F. Psychometric properties of the Turkish version of the fear avoidance components scale in patients with chronic pain related to musculoskeletal disorders. Wiener klinische Wochenschrift 2023:1-7.

[193] Turk DC, Robinson JP, Sherman JJ, Burwinkle T, Swanson K. Assessing fear in patients with cervical pain: development and validation of the Pictorial Fear of Activity Scale-Cervical (PFActS-C). Pain 2008;139(1):55-62.

[194] Van Breukelen GJ, Vlaeyen JW. Norming clinical questionnaires with multiple regression: the Pain Cognition List. Psychol Assess 2005;17(3):336-344.

[195] Van Damme S, Van Ryckeghem DML, Wyffels F, Van Hulle L, Crombez G. No pain no gain? Pursuing a competing goal inhibits avoidance behavior. Pain 2012;153(4):800-804.

[196] Van Ryckeghem D, Crombez G. Assessment and measurement in health psychology. Comprehensive clinical psychology, vol 8: health psychology: Elsevier, 2022. pp. 85-94.

[197] Vasiliou VS, Michaelides MP, Kasinopoulos O, Karekla M. Psychological Inflexibility in Pain Scale: Greek adaptation, psychometric properties, and invariance testing across three pain samples. Psychological Assessment 2019;31(7):895.

[198] Visscher CM, Ohrbach R, van Wijk AJ, Wilkosz M, Naeije M. The tampa scale for kinesiophobia for temporomandibular disorders (TSK-TMD). Pain 2010;150(3):492-500.

[199] Vlaeyen JW, Kole-Snijders AM, Boeren RG, Van Eek H. Fear of movement/(re) injury in chronic low back pain and its relation to behavioral performance. Pain 1995;62(3):363-372.

[200] Vlaeyen JW, Kole-Snijders AM, Rotteveel AM, Ruesink R, Heuts PH. The role of fear of movement/(re) injury in pain disability. Journal of occupational rehabilitation 1995;5:235-252.

[201] Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000;85(3):317-332.

[202] Vlaeyen JW, Linton SJ. Fear-avoidance model of chronic musculoskeletal pain: 12 years on. Pain 2012;153(6):1144-1147.

[203] Vlaeyen JWS, Morley S. Active despite pain: the putative role of stop-rules and current mood. Pain 2004;110(3):512-516.

[204] Volders S, Boddez Y, De Peuter S, Meulders A, Vlaeyen JW. Avoidance behavior in chronic pain research: a cold case revisited. Behaviour research and therapy 2015;64:31-37.

[205] Waardenburg S, Visseren L, van Daal E, Brouwer B, van Zundert J, van Kuijk SMJ, Lousberg R, Jongen EMM, Leue C, de Meij N. Do Men and Women Have a Different Association between Fear-Avoidance and Pain Intensity in Chronic Pain? An Experience Sampling Method Cohort-Study. J Clin Med 2022;11(19).

[206] Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993;52(2):157-168.

[207] Wei X, Xu X, Zhao Y, Hu W, Bai Y, Li M. The Chinese version of the Tampa Scale for Kinesiophobia was cross-culturally adapted and validated in patients with low back pain. Journal of clinical epidemiology 2015;68(10):1205-1212.

[208] Wicksell RK, Lekander M, Sorjonen K, Olsson GL. The Psychological Inflexibility in Pain Scale (PIPS)–statistical properties and model fit of an instrument to assess change processes in pain related disability. European Journal of Pain 2010;14(7):771. e771-771. e714.

[209] Wicksell RK, Renöfält J, Olsson GL, Bond FW, Melin L. Avoidance and cognitive fusion–central components in pain related disability? Development and preliminary validation of the Psychological Inflexibility in Pain Scale (PIPS). European Journal of Pain 2008;12(4):491-500.

[210] Woby SR, Roach NK, Urmston M, Watson PJ. Psychometric properties of the TSK-11: a shortened version of the Tampa Scale for Kinesiophobia. Pain 2005;117(1-2):137-144.

[211] Wong WS, Jensen MP, Mak KH, Tam BK, Fielding R. Preliminary psychometric properties of the Chinese version of the Chronic Pain Coping Inventory (ChCPCI) in a Hong Kong Chinese population. The Journal of Pain 2010;11(7):672-680.

[212] Wong WS, Kwok HY, Luk K, Chow YF, Mak KH, Tam B, Wong ET, Fielding R. Fear of movement/(re) injury in Chinese patients with chronic pain: Factorial validity of the Chinese version of the Tampa Scale for Kinesiophobia. Journal of Rehabilitation Medicine 2010;42(7):620-629.

[213] Zhou X-Y, Xu X-M, Wang F, Wu S-Y, Yang Y-L, Li M, Huang J-M, Wei X-Z. Validations and psychological properties of a simplified Chinese version of pain anxiety symptoms scale (SC-PASS). Medicine 2017;96(10):e5626.

Figures

Figure 1. Flow chart illustrating the selection of articles for the review according to the PRISMA guidelines [124].

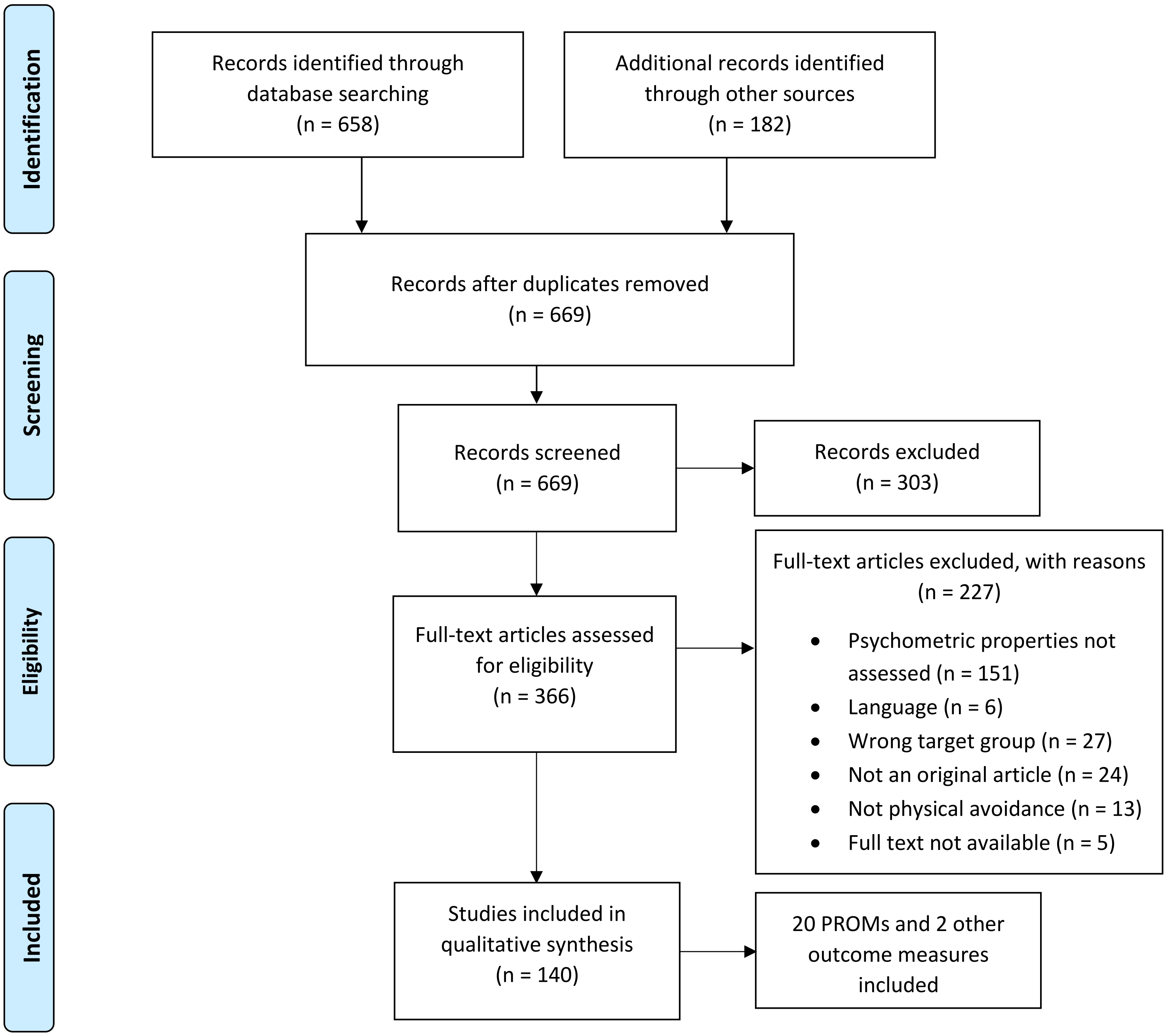
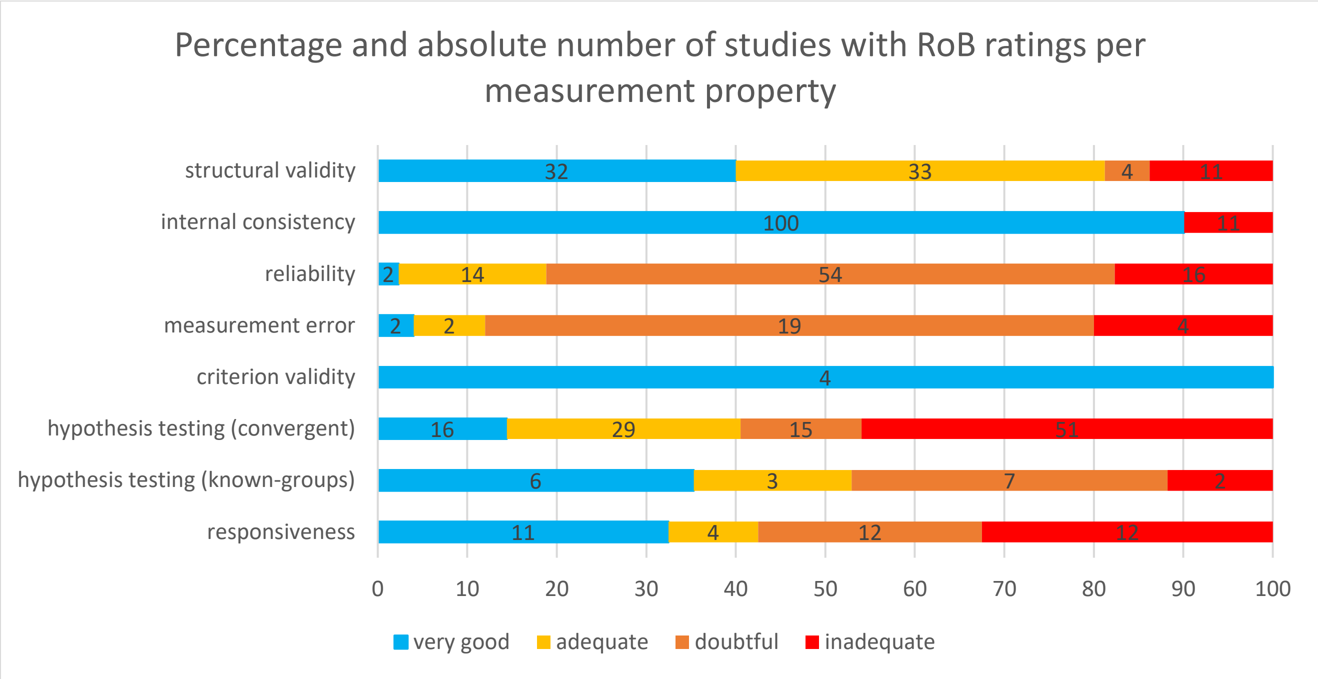


Figure 2. Percentage and absolute number of studies with the different Risk of Bias ratings per measurement property.



Tables

Table 1. Characteristics of patient-reported outcome measures (PROMs).

PROM* [development study]	Target population	Mode of administration	Recall period	Subscales (number of items)	Response options	Original language	Available translations
ADAP [8]	persons with shoulder pain	Rating of photographs of activities	now	free movement, high effort, self-care (15 items)	0 (not avoid at all) – 10 (avoid extremely)	Brazilian Portuguese	
AEQ [68]	persons with LBP	Self-report	undefined	anxiety/depression, help-/hopelessness, catastrophizing, avoidance of social activities scale, avoidance of physical activities scale, positive mood, thought suppression, behavioral endurance scale with two sub-scales humor/distraction scale and pain persistence scale (49 items)	10 affective items & 16 cognitive items: 0 (never) – 6 (always), 23 behavioral items: 0 (never) – 5 (always)	German	Korean, Persian, Spanish
APS [51]	chronic pain patients	Self-report	undefined	pain avoidance, activity avoidance, task-contingent persistence, excessive persistence, pain-contingent persistence, pacing to increase activity levels, pacing to conserve energy for valued activities, and pacing to reduce pain (24 items)	0 (not at all) - 4 (always)	Spanish	
BSFAQ [103]	burn patients	Self-report	now	(5 items)	0 (strongly disagree) – 3 (strongly agree)	English, French	
CAS D-65+ [157]	elderly (65+) with LBP	Self-report	undefined	catastrophizing, avoidance (11 items)	0 (never) – 5 (always)	German	

CPCI [78]	chronic pain patients	Self-report	past week	guarding, resting, asking for assistance, medication use relaxation, task persistence, exercise/stretch, coping self-statements, seeking social support (65 items)	number of days: 0-7	English	Brazilian Portuguese
CPCI-42 [166]	chronic pain patients	Self-report	past week	guarding, resting, asking for assistance, relaxation, task persistence, exercise/stretch, coping self-statements, seeking social support, opioid medication use, non-steroidal medication use, sedative-hypnotic medication use (42 items)	number of days: 0-7	English	Chinese, Italian, Korean, Polish, Spanish
FABQ [206]	persons with LBP	Self-report	undefined	fear-avoidance beliefs about work, fear-avoidance beliefs about physical activity (16 items)	0 (completely disagree) – 6 (completely agree)	English	Arabic, Brazilian Portuguese, Chinese, Danish, Dutch, Finnish, French, German, Greek, Gujarati, Hausa, Hindi, Igbo, Italian, Japanese, Kannada, Norwegian, Persian, Spanish, Swedish, Tamil, Turkish, Yoruba
FACS [135]	chronic musculoskeletal pain	Self-report	last week	(20 items)	0 (completely disagree) – 5 (completely agree)	English	Brazilian Portuguese, Dutch, French, Gujarati, Mandarin, Persian, Serbian, Spanish (European & Central American), Turkish
NRP [79]	general population	Self-report	undefined	NRP scales: despondent, fear, avoidant (12 items) PR scales: happy/hopeful, approach (8 items)	0 (I never feel this way) – 4 (I feel this all the time)	English	

PARQ [114]	chronic pain patients	Self-report	undefined	avoidance, pacing, confronting (21 items)	0 (never) – 5 (always)	English	
PASS [116]	chronic pain patients	Self-report	undefined	cognitive, escape and avoidance, fear, physiological anxiety (40 items)	0 (never) – 5 (always)	English	
PASS-20 [113]	chronic pain patients	Self-report	undefined	cognitive, escape and avoidance, fear, physiological anxiety (20 items)	0 (never) – 5 (always)	English	Arabic, Chinese, Dutch, German, Korean, Persian, Spanish
PCI [92]	chronic pain patients	Self-report	undefined	pain transformation; distraction; reducing demands; retreating; worrying; resting (33 items)	1 (hardly ever) - 4 (very often)	Dutch	
PIPS [209]	chronic pain patients	Self-report	now	avoidance, cognitive fusion (16 items)	1 (never true) – 7 (always true)	Swedish	Dutch, English, German, Greek, Japanese, Persian, Spanish
POAM-P [22]	chronic pain patients	Self-report	undefined	avoidance, overdoing, pacing (30 items)	0 (not at all) – 4 (all the time)	English	French, Japanese, Spanish, Turkish
TSK [121]	chronic pain patients	Self-report	undefined	somatic focus, activity avoidance (17 items)	1 (strongly disagree) – 4 (strongly agree)	English	Arabic, Brazilian Portuguese, Cantonese, Chinese, Dutch, Finnish, French, German, Greek, Gujarati, Iranian, Italian, Japanese, Norwegian, Persian, Portuguese, Spanish, Swedish, Thai
TSK-11 [210]	chronic pain patients	Self-report	undefined	somatic focus, activity avoidance (11 items)	1 (strongly disagree) – 4 (strongly agree)	English	Brazilian Portuguese, Chinese, Japanese, Malay, Marathi, Spanish, Swedish, Thai

TSK-13 [32]	chronic pain patients	Self-report	undefined	somatic focus, activity avoidance (13 items)	1 (strongly disagree) – 4 (strongly agree)	English	
TSK-TMD [198]	temporomandibular pain patients	Self-report	undefined	somatic focus, activity avoidance (12 items)	1 (strongly disagree) – 4 (strongly agree)	Dutch	Brazilian Portuguese, English, French, Korean, Mandarin, Spanish

* Each version of a PROM is considered a separate PROM.

Table 2. Characteristics of other measurement instruments.

Instrument	Target population	Mode of administration	Rating	Original language	Available translations
BAT-Back [72]	persons with CLBP	physical test	0 (no avoidance/safety behavior), 1 (safety behavior), 2 (avoidance of movement)	German	Turkish
Clinician-reported fear-avoidance beliefs [21]	persons with LBP	therapist assessment	0 (no fear-avoidance) – 10 (very high fear-avoidance)	English	

Table 3. COSMIN definitions and criteria for good measurement properties of measurement instruments.

Domain	Measurement property	Definition	Rating	Criteria
Reliability	Internal consistency*	The degree of the interrelatedness among the items	+	At least low evidence[a] for sufficient structural validity[b] AND Cronbach's alpha(s) ≥ 0.70 for each unidimensional scale or subscale[c]
			?	Criteria for "At least low evidence[a] for sufficient structural validity[b]" not met
			-	At least low evidence[a] for sufficient structural validity[b] AND Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale[c]
	Reliability	The proportion of the total variance in the measurements which is due to 'true' differences between participants	+	ICC or weighted Kappa ≥ 0.70
			?	ICC or weighted Kappa not reported
			-	ICC or weighted Kappa < 0.70
	Measurement error	The systematic and random error of a participant's score that is not attributed to true changes in the construct to be measured	+	SDC or LoA < MIC[b]
			?	MIC not defined
			-	SDC or LoA > MIC[b]
Validity	Content validity	The degree to which the content of an instrument is an adequate reflection of the construct to be measured		No objective criteria
	Structural validity*	The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured	+	**CTT** CFA: CFI or TLI or comparable measure > 0.95 OR RMSEA < 0.06 OR SRMR < 0.08[d]

			IRT/Rasch No violation of <u>unidimensionality</u>[e]: CFI or TLI or comparable measure > 0.95 OR RMSEA < 0.06 OR SRMR < 0.08 AND no violation of <u>local independence</u>: residual correlations among the items after controlling for the dominant factor < 0.20 OR Q3's < 0.37 AND no violation of <u>monotonicity</u>: adequate looking graphs OR item scalability > 0.30 AND adequate <u>model fit</u> IRT: $\chi^2 > 0.001$ Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardized values > −2 and < 2
		?	CTT: not all information for '+' reported IRT/Rasch: model fit not reported
		-	Criteria for '+' not met
Hypothesis testing for construct validity	The degree to which the scores of an instrument are consistent with hypotheses (based on the assumption that the instrument validly measures the construct to be measured)	+	The result is in accordance with the hypothesis[f]
		?	No hypothesis defined (by the review team)
		-	The result is not in accordance with the hypothesis[f]
		+	No important differences found between group factors (such as age,

Cross-cultural validity/ measurement invariance*	The degree to which the performance of the items on a translated instrument or a version adapted to a different culture or patient group are an adequate reflection of the performance of the items of the original version of the instrument		gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's $R^2 < 0.02$)
		?	No multiple group factor analysis OR DIF analysis performed
		-	Important differences between group factors OR DIF was found
Criterion validity*	The degree to which the scores of an instrument are an adequate reflection of a 'gold standard'	+	Correlation with gold standard ≥ 0.70 OR AUC ≥ 0.70
		?	Not all information for '+' reported
		-	Correlation with gold standard < 0.70 OR AUC < 0.70
Responsiveness	The ability of an instrument to detect change over time in the construct to be measured	+	The result is in accordance with the hypothesis[f] OR AUC ≥ 0.70
		?	No hypothesis defined (by the review team)
		-	The result is not in accordance with the hypothesis[f] OR AUC < 0.70

The criteria are based on [186], [126] and Prinsen et al. [151; 152]

AUC area under the curve, *CFA* confirmatory factor analysis, *CFI* comparative fit index, *CTT* classical test theory, *DIF* differential item functioning, *ICC* intraclass correlation coefficient, *IRT* item response theory, *LoA* limits of agreement, *MIC* minimal important change, *RMSEA* root mean square error of approximation, *SEM* standard error of measurement, *SDC* smallest detectable change, *SRMR* standardized root mean residuals, *TLI* Tucker–Lewis index

"+" = sufficient, "–" = insufficient, "?" = indeterminate

* relevant for patient-reported outcome measures (PROMs) but not for performance-based outcome measures.

[a] As defined by grading the evidence according to the GRADE approach; [b] This evidence may come from different studies; [c] The criteria 'Cronbach alpha < 0.95' was deleted, as this is relevant in the development phase of a PROM and not when evaluating an existing PROM; [d] To rate the quality of the summary score, the factor structures should be equal across studies; [e] Unidimensionality refers to a factor analysis per subscale, while structural validity refers to a factor analysis of a (multidimensional) patient-reported outcome measure; [f] The results of all studies should be taken together and it should then be decided if 75% of the results are in accordance with the hypotheses.

Table 4. Characteristics of included studies.

Instrument	Author (year)	Sample Characteristics			Country	Language	Psychometric properties
		N	Demographics	Pain type			
ADAP	Ansanello 2022 [8]	156	mean age = 47.8, SD = 17.2; 59% female	shoulder pain	Brazil	Brazilian Portuguese	content validity, structural validity, internal consistency
	Ansanello 2023 [9]	100	mean age = 44.9, SD = 15.9; 57% female	shoulder pain	Brazil		reliability, measurement error, construct validity
AEQ	Hasenbring 2009 [68]	191	mean age = 44.96, SD = 11.3; 55% female	LBP	Germany	German	structural validity, internal consistency, construct validity
	Abad 2020 [2]	123	mean age = 34.85; SD= 11.29; 76% female	non-specific neck pain	Iran	Persian	internal consistency, reliability, measurement error, construct validity
	Ruiz-Párraga 2015 [170]	150	mean age = 48.27, SD = 6.7; 62% female	musculoskeletal	Spain	Spanish	structural validity, internal consistency, construct validity
APS	Esteve 2016 [51]	402	unclear, different samples mixed	fibromyalgia, rheumatic diseases	Spain	Spanish	content validity, structural validity, internal consistency, construct validity
BAT-Back	Holzapfel 2016 [72]	128	64.9% female	LBP vs HC	Germany	German	internal consistency, reliability, construct validity
	Küçükakkaş 2020 [100]	155	mean age = 44.4, SD = 13.2; 48.7% female	LBP vs HC	Turkey	Turkish	structural validity, internal consistency, reliability, construct validity
BSFAQ	Chen 2023 [24]	31	mean age = 50.1, SD = 14.8; 25.5% female	burns	Canada	French	construct validity

CAS-D-65+	Quint 2011 [157]	68	mean age = 74.1; 76.5% female	back pain	Germany	German	internal consistency, reliability, construct validity
Clinician-rep. FAB	Calley 2010 [21]	80	mean age = 46.6, SD = 11.5; 57.5% female	LBP	US	English	reliability, construct validity
CPCI	Hadjistavropoulos 1999 [66]	210	mean age = 40, SD = 10.4; 39% female	mixed	Canada	English	structural validity, construct validity
	Jensen 1995 [78]	254	unclear, different samples mixed	mixed	US		internal consistency, reliability, construct validity
	Tan 2005 [184]	564	mean age = 50.8, SD = 11.4; 9.7% female	mixed	US		structural validity, construct validity
	Souza 2018 [179]	59	66% female	mixed	Brazil	Brazilian Portuguese	content validity, internal consistency
	Souza 2021 [180]	705	mean age = 53.8, SD = 14.26; 68.4% female	nonspecific, mixed	Brazil		structural validity, internal consistency
	Truchon 2006 [191]	827	mean age = 44, SD = 12.6; 70% female	mixed	Canada, France	French	structural validity, internal consistency
CPCI, CPCI-42	Romano 2003 [166]	154	mean age = 43.46, 51% female	mixed	US	English	internal consistency, reliability, criterion validity, construct validity, responsiveness
CPCI-42	Wong 2010a [211]	208	mean age = 40.95, SD = 11.28; 54.3% female	musculoskeletal	China	Chinese	content validity, structural validity, internal consistency, construct validity
	Monticone 2013 [130]	270	mean age = 55.4, SD = 14.6; 55.9% female	non-specific (musculoskeletal)	Italy	Italian	content validity, structural validity, internal consistency, reliability, construct validity

	Ko 2010 [88]	142	mean age = 47.9, SD = 11.9; 73% female	LBP	Korea	Korean	content validity, structural validity, internal consistency, reliability, construct validity
	Misterska 2014 [123]	90	mean age = 43.47, SD = 10.21; 60% female	LBP	Poland	Polish	content validity, internal consistency, reliability, construct validity
	Garcia-Campayo 2007 [55]	402	mean age = 45.3, SD = 6.8; 89.8% female	fibromyalgia	Spain	Spanish	structural validity, internal consistency, reliability, construct validity
FABQ	Cleland 2008a [33]	263	51.8% female	LBP	US	English	construct validity
	Riley 2019 [161]	30	mean age = 48.1, SD = 15.6; 24 female	shoulder pain			structural validity, construct validity, responsiveness
	Inrig 2012 [75]	187	mean age = 45.2, SD = 9.68; 46% female	upper extremity	Canada		internal consistency, reliability, measurement error, construct validity
	Waddell 1993 [206]	184	mean age = 39.7, SD = 11.7; 44.3% female	LBP and/or sciatica	UK		structural validity, internal consistency, reliability, construct validity
	Laufer 2012 [105]	113	unclear, different samples mixed	LBP	Israel	Arabic	content validity, internal consistency, reliability, construct validity
	Alanazi 2017 [5]	110	mean age = 47.3, 35.2% female	LBP	Saudi Arabia		reliability, construct validity
	Salama 2020 [173]	80	mean age = 47, SD = 13.4; 67.5% female	LBP	Egypt		construct validity
	Cheung 2018 [25]	100	mean age = 57, SD = 12.5; 57% female	LBP	China	Chinese	content validity, internal consistency, construct validity

Pei 2010 [142]	245	mean age = 47.5, SD = 14.1, 63% female	LBP			content validity, structural validity, internal consistency, reliability, construct validity, responsiveness
Lee 2006 [106]	476	unclear, different samples mixed	neck pain	Hong Kong		content validity, structural validity, internal consistency, reliability, construct validity, responsiveness
Trolle 2019 [189]	45	mean age = 48.1, SD = 15; 40% female	shoulder impingement syndrome	Denmark	Danish	reliability, measurement error, construct validity, responsiveness
Sørensen 2021 [183]	52	mean age = 57.4, SD = 10.1; 50% female	shoulder pain			responsiveness
Terho 2016 [185]	66	mean age = 45.8 SD = 12.9; 57.6% female	LBP	Finland	Finnish	content validity, structural validity, internal consistency, reliability
Chaory 2004 [23]	244	mean ages = 46.0, SD = 7.5 / 45.3, SD = 9.2 /42.5 = 7.5 (range 26–54)	LBP	France	French	structural validity, reliability, measurement error, construct validity, responsiveness
Pagels 2023 [139]	49	mean age = 41.8, SD = 12.8; 48.98% female	shoulder pain	Germany	German	internal consistency, construct validity
Pfingsten 1997 [146]	87	mean age = 40.9, SD = 9.6; 57.5% female	LBP			structural validity, internal consistency, reliability, construct validity
Pfingsten 2000 [145]	302	mean age = 44.6, SD = 10.6; 47.7% female	LBP			structural validity, internal consistency, reliability, construct validity
Pfingsten 2004 [144]	302	mean age: 44.6, SD = 10.6; 47.7% female	back pain			internal consistency, construct validity

Staerkle 2004 [182]	255	mean age = 56.9, SD = 15.5; 55.69% female	LBP			structural validity, internal consistency, reliability, construct validity
Georgoudis 2007 [60]	70	mean age = 42.2, SD = 12; 82.9% female	LBP	Greece	Greek	content validity, structural validity, internal consistency, reliability, construct validity, responsiveness
Bid 2016 [18]	30	mean age = 41.8, SD = 11.36; 63.3% female	LBP	India	Gujarati	content validity, internal consistency, reliability, measurement error, construct validity
Bid 2019 [15]	128	mean age = 41.23, SD = 7.48; 60.9% female	LBP	India	Gujarati	structural validity, construct validity, responsiveness
Ibrahim 2019 [73]	200	unclear, different samples mixed	LBP	Nigeria	Hausa	content validity, structural validity, internal consistency, reliability, measurement error, construct validity, responsiveness
Kaka 2015 [80]	54	mean age = 39.3, SD = 10.86; 33.3% female	neck pain			content validity, reliability, criterion validity, construct validity
Igwesi-Chidobe 2019 [74]	50	mean age = 45.2, SD = 11.55; 64% female	LBP		Igbo	content validity, structural validity, internal consistency, reliability, measurement error, construct validity
Panhale 2019 [140]	100	mean age = 36.89, SD = 7.78	LBP	India	Hindi	content validity, structural validity, internal consistency, reliability, construct validity
Franchignoni 2020 [52]	155	mean age = 43, SD = 11; 43% female	LBP	Italy	Italian	structural validity, internal consistency
Monticone 2020 [131]	129	mean age = 48, SD = 16; 66.66% female	LBP			internal consistency, reliability, measurement error, responsiveness

Monticone 2012 [129]	180	mean age = 44.1, SD = 11.3; 43% female	LBP			content validity, structural validity, internal consistency, reliability, measurement error, construct validity
Pruneti 2014 [154]	250	mean age = 41.84, SD = 11.06; 48% female	back pain			structural validity, internal consistency, reliability
Matsudaira 2014 [111]	178 6	mean age = 48.7, 49.6% female	LBP	Japan	Japanese	structural validity, internal consistency, construct validity
Kumar 2020 [99]	60	mean age = 41.5, SD = 10.59	LBP	India	Kannada	content validity, internal consistency, reliability
Panhale 2018 [141]	100	mean age = 33.97, SD = 11.54; 45% female	LBP	India	Marathi	content validity, internal consistency, reliability, construct validity
Grotle 2006 [64]	173	unclear, different samples mixed	LBP	Norway	Norwegian	structural validity, internal consistency, reliability, measurement error, construct validity, responsiveness
Rostami 2014 [169]	136	mean age = 48.7, SD = 13.13; 47.7% female	LBP	Iran	Persian	structural validity, internal consistency, reliability, construct validity
Kovacs 2006 [91]	156	mean age = 45.7; 60.9% female	LBP	Spain	Spanish	content validity, internal consistency, reliability
Dayalan 2022 [42]	50	mean age = 39.9, SD = 15.5; 60% female	LBP	India	Tamil	internal consistency, reliability, construct validity
Korkmaz 2009 [90]	150	mean age = 44.7, SD = 13.06; 64% female	LBP, leg	Turkey	Turkish	structural validity, internal consistency, reliability, construct validity, responsiveness

	Namli 2022 [134]	195	mean age = 62.67, SD = 4.57; 79% female	knee osteoarthritis			internal consistency, construct validity
	Ozuberk 2023 [138]	175	mean age = 43.41, SD = 10.96; 79.4% female	neck pain			internal consistency, reliability, construct validity
	Mbada 2021 [112]	131	mean age = 53.6, SD = 11.6; 45.8% female	LBP	Nigeria	Yoruba	content validity, structural validity, internal consistency, reliability, measurement error, construct validity
FABQ, PASS, TSK	Crombez 1999 [38]	104	unclear, different samples mixed	LBP	Belgium	Dutch	internal consistency, construct validity
FABQ, TSK	Cleland 2008b [34]	78	mean age = 42, SD = 11.3; 68% female	neck pain	US	English	reliability, construct validity
FABQ, TSK	de Souza 2008 [44]	50	mean age = 45.9, SD = 12; 54% female	LBP	Brazil	Brazilian Portuguese	internal consistency, reliability, construct validity, responsiveness
FABQ, TSK	Askary-Ashtiani 2014 [12]	166	45% female	neck pain	Iran	Persian	internal consistency, reliability, construct validity
FABQ, TSK	Saadat 2023 [172]	100	mean = 42.01, SD = 10.18; 69% female	neck pain	Iran	Persian	responsiveness
FABQ, TSK	Dedering 2013 [45]	41	60.97% female	cervical radiculopathy	Sweden	Swedish	internal consistency, reliability, construct validity
FABQ, TSK-11	George 2010 [59]	53	mean age = 44.3, SD = 18.5; 80% female	LBP	US	English	reliability, measurement error, construct validity
FABQ, TSK-11	Kamonseki 2021 [81]	178	mean age = 39.7, SD = 14.01; 41.6% female	shoulder pain	Brazil	Brazilian Portuguese	structural validity, internal consistency, reliability, measurement error, responsiveness

FABQ, TSK-11	Mintken 2010 [122]	80	mean age = 41.2, SD = 13.2, 60% female	shoulder pain	US	English	structural validity, internal consistency, reliability, construct validity
FACS	Neblett 2016 [135]	788	unclear	musculoskeletal	US	English	content validity, internal consistency, reliability, construct validity
	Neblett 2017 [136]	426	mean age = 47.1, SD = 10.9; 39% female	musculoskeletal			structural validity, responsiveness
	De Baets 2023 [43]	224	mean age = 48.6, SD = 16; 66.07% female	musculoskeletal pain	Belgium	Dutch	structural validity, internal consistency, reliability, measurement error, construct validity
	Duport 2023 [46]	55	mean age = 51.15, SD = 16.47; 45.5% female	musculoskeletal pain	France	French	internal consistency, reliability, construct validity
	Bid 2020 [16]	150	mean = 47.36, SD = 9.85; 68.67% female	musculoskeletal pain	India	Gujarati	content validity, internal consistency, reliability, measurement error, construct validity
	Knezevic 2018 [87]	322	mean age = 52.98, SD = 12.33; 67.1% female	musculoskeletal pain	Serbia	Serbian	structural validity, internal consistency, reliability, measurement error, construct validity
	Cuesta-Vargas 2020 [40]	330	mean age = 55.04, SD = 12.7; 45.2% female	musculoskeletal pain disorders	Spain	Spanish	structural validity, internal consistency, construct validity
	González Aroca 2023 [62]	208	mean age = 48.45, SD = 6.05; 52.9% female	unilateral subacromial shoulder pain			construct validity
	Turan 2023 [192]	208	mean age = 46.2, SD = 11.4; 55.8% female	musculoskeletal pain	Turkey	Turkish	structural validity, internal consistency, reliability, construct validity

NRP	Jensen 2017 [79]	395	75.95% female	various	US	English	structural validity, internal consistency, reliability, construct validity
PARQ	McCracken 2007 [114]	276	mean age = 46.6, SD = 13.7; 65.6% female	mixed	UK	English	structural validity, internal consistency, construct validity
PASS	McCracken 1992 [116]	104	mean age = 45, SD = 13.4, 53.8% female	mixed	US		internal consistency, construct validity
PASS-20	Coons 2004 [35]	201	mean age = 41.5,SD =15.5, 55% female	mixed	Canada	English	structural validity, internal consistency, reliability, construct validity
	McCracken 2002 [113]	282	mean age = 46.5, SD = 13.8; 66% female	mixed	UK		internal consistency, criterion validity, construct validity
	Kreddig 2015 [94]	195	mean age = 42.45, SD = 11.29; 55% female	LBP	Germany	German	structural validity, internal consistency, construct validity
	Cho 2010 [27]	166	mean age = 48.7, SD = 13.04; 70.5% female	mixed	Korea	Korean	structural validity, internal consistency, reliability, construct validity
	Zhou 2017 [213]	249	mean age = 58.4, SD = 13.4; 66.7% female	?	China	Chinese	content validity, structural validity, internal consistency, reliability, construct validity
PCI	Kraaimaat 2003 [93]	1055	unclear, different samples mixed	mixed	Dutch	Dutch	internal consistency, reliability, construct validity
PIPS	Trompetter 2014 [190]	428	mean age = 43.7, SD = 12.5; 72.2% female	mixed, musculoskeletal	the Netherlands	Dutch	structural validity, internal consistency, construct validity
	Barke 2015 [13]	182	mean age = 51.0, SD = 10.5; 70.3% female	back pain	Germany	German	content validity, structural validity, internal consistency, construct validity

	Vasiliou 2019 [197]	394	unclear, different samples mixed	musculoskeletal and other (incl. headache)	Greece	Greek	structural validity, internal consistency, reliability, construct validity
	Nagasawa 2021 [133]	120	mean age = 73.8, SD = 7.8; 61.7% female	LBP, knee pain	Japan	Japanese	structural validity, internal consistency, reliability, construct validity
	Rodero 2013 [162]	259	mean age = 52.4, SD = 8.5; 95.6% female	fibromyalgia	Spain	Spanish	structural validity, internal consistency, reliability, construct validity
	Wicksell 2008 [209]	203	mean age = 45.5, SD = 10.15; 80.8% female	fibromyalgia, migraine, whiplash, LBP	Sweden	Swedish	structural validity, internal consistency, construct validity
	Wicksell 2010 [208]	611	mean age = 49, SD = 12.8; 74.8% female	whiplash			structural validity, internal consistency, construct validity
POAM-P	Cane 2013 [22]	559	unclear, different samples mixed	mixed	Canada	English	internal consistency, construct validity, responsiveness
	Benaim 2017 [14]	595	mean age = 43, SD = 12	musculoskeletal	Switzerland	French	content validity, internal consistency, reliability, construct validity
TSK	French 2007 [53]	200	mean age = 40, SD = 10.6; 54% female	back / neck	Canada	English	reliability
	Alanazi 2021 [6]	82	mean age = 29.7; SD= 9.6; 29.3% female	LBP	Saudi Arabia	Arabic	internal consistency, reliability, measurement error, construct validity
	Siqueira 2007 [177]	50	mean age = 41.98, SD = 13.76; 76% female	LBP	Brazil	Brazilian Portuguese	structural validity, internal consistency
	Wong 2010b [212]	325	mean age = 39.72, SD = 13.88; 52.3% female	any	China	Cantonese	structural validity, internal consistency, construct validity

Lamé 2008 [102]	50	mean age = 54.7, SD = 13.1; 60% female	mixed			internal consistency, reliability
Pulles 2020 [155]	359	mean age = 45.28, SD = 11.02; 66.3% female	musculoskeletal pain			measurement error, construct validity, responsiveness
Roelofs 2004 [163]	616	unclear, different samples mixed	LBP, fibromyalgia			structural validity, internal consistency, construct validity
Vlaeyen 1995a [199]	103	mean age = 39.0/42.9, SD = 7.7/8.9; 56.3% female (only S1)	LBP			internal consistency, construct validity
Vlaeyen 1995b [200]	162	unclear, different samples mixed	LBP			structural validity, internal consistency, construct validity
Koho 2014 [89]	94	mean age = 47, SD= 8; 58.5% female	musculoskeletal pain	Finland	Finnish	internal consistency, reliability
Georgoudis 2022 [61]	70	mean age = 42.2, SD = 12; 82.9% female	LBP	Greece	Greek	internal consistency, reliability, construct validity
Bid 2018 [17]	160	mean age = 43.26, SD = 14.51; 67.5% female	LBP	India	Gujarati	content validity, structural validity, internal consistency, reliability, construct validity
Monticone 2016a [127]	180	mean age = 55.4, SD = 10.5; 59.44% female	LBP	Italy	Italian	measurement error, responsiveness
Monticone 2016b [128]	205	mean age = 50.9, SD = 9.1; 58.54% female	LBP			measurement error, responsiveness
Monticone 2010 [132]	178	mean age = 61.5, SD = 13.2; 70.2% female	LBP			internal consistency, construct validity

	Haugen 2008 [69]	466	mean age = 43.6, SD = 11.5; 42.5% female	sciatica	Norway	Norwegian	structural validity, internal consistency, reliability, construct validity, responsiveness
	Lundberg 2004 [110]	102	mean age = 45.3, SD = 12.5; 50.98% female	LBP	Sweden	Swedish	structural validity, internal consistency, reliability, construct validity
	Areeudomwong 2017 [10]	80	mean age = 65.53, SD= 9.39; 66.25% female	knee pain	Thailand	Thai	internal consistency, reliability, construct validity
TSK, TSK-11	Woby 2005 [210]	111	mean age = 43.4, SD = 10.5; 48.65% female	LBP	UK	English	content validity, structural validity, internal consistency, reliability, construct validity, responsiveness
TSK, TSK-11	Roelofs 2007[a] [164]	3934	mean age = 41.7, SD = 8.7; 67% female	work-related upper extremity disorders	the Netherlands	Dutch	structural validity, internal consistency, construct validity
TSK, TSK-11	Kikuchi 2015 [86]	956	mean age = 45.5, SD = 10.4; 29% female	whiplash & LBP	Japan	Japanese	content validity, reliability, measurement error, construct validity
TSK, TSK-11, TSK-13	Eiger 2023 [48]	77	mean age = 49.9, SD = 14.4; 66.2% female	any	Denmark	Danish	reliability, measurement error
TSK, TSK-13	Burwinkle 2015 [19]	233	mean age = 43.79, SD = 10.83; 100% female	fibromyalgia	US	English	structural validity, internal consistency
TSK, TSK-13	Goubert 2004 [63]	277	mean age = 41.33, SD = 10.9; 66.4% female	LBP, FM	Belgium, the Netherlands	Dutch	structural validity, internal consistency
TSK, TSK-13	Heuts 2004 [71]	254	mean age = 51.7, SD = 5	osteoarthritis	the Netherlands		structural validity, internal consistency

TSK, TSK-13	Wei 2015 [207]	142	mean age = 52.3, SD = 10.2; 47.9% female	LBP		Chinese	structural validity, internal consistency, reliability, construct validity
TSK-11	Hapidou 2012 [67]	74	mean age = 43.8, SD = 9.3; 61% female	any	Canada	English	reliability, measurement error, construct validity, responsiveness
	Santo Salvador 2021 [174]	130	mean age = 45.5, SD = 11.1; 98.5% female	fibromyalgia	Brazil	Brazilian Portuguese	internal consistency, reliability, measurement error, criterion validity, construct validity
	Cai 2019 [20]	254	mean age = 63.1, SD = 9.8; 56.3% female	total knee athroplasty	China	Chinese	structural validity, internal consistency, reliability, construct validity
	Rusu 2014 [171]	191	mean age = 50.1, SD = 11.3; 55% female	LBP	Germany	German	content validity, structural validity, internal consistency, construct validity
	Satpute 2019 [175]	100	mean age = 38.9, SD = 11.34; 47.9% female	LBP	India	Marathi	content validity, reliability, construct validity
	Gómez-Perez 2011 [65]	211	unclear, different samples mixed	mixed	Spain	Spanish	content validity, structural validity, internal consistency, reliability, construct validity
	Larsson 2014 [104]	433	mean age = 74.8, SD = 7.5; 63.5% female	mixed	Sweden	Swedish	structural validity, internal consistency, construct validity
TSK-11, TSK-13	Tkachuk 2012 [188]	276	mean age = 47.76, SD= 12.35; 65% female	mixed	Canada	English	internal consistency, reliability, measurement error, responsiveness
TSK-13	Geisser 2000 [56]	133	mean age = 41.7, SD = 8.5; 43.61% female	back pain	US	English	structural validity, construct validity

	Cordeiro 2013 [36]	166	mean age = 50.55, SD = 10.80; 63.3% female	LBP	Portugal	Portuguese	structural validity, internal consistency, reliability, construct validity, responsiveness
TSK-TMD	Aguiar 2017 [4]	100	mean age = 36.88, SD = 9.8; 100% female	temporo-mandibular pain	Brazil	Brazilian Portuguese	content validity, structural validity, internal consistency, reliability, measurement error, construct validity
	He 2016 [70]	160	mean age = 45.2, SD = 15.8; 54.4% female	temporo-mandibular pain	China	Chinese	content validity, internal consistency, reliability, construct validity
	Visscher 2010 [198]	301	mean age = 41.3, SD = 14.1; 81% female	temporo-mandibular pain	the Netherlands	Dutch	structural validity, internal consistency, reliability, construct validity
	LaTouche 2020 [101]	125	mean age = 45.58, SD = 12.92; 67.3% female	temporo-mandibular pain	Spain	Spanish	content validity, structural validity, internal consistency, reliability, measurement error, construct validity

[a]The study by Roelofs et al.[164] contained a sub-study in which several datasets (including multiple language versions of the TSK), some of which are part of this review, were pooled and re-analysed. Therefore, this sub-study was not considered here.

responsiveness = responsiveness to change.

Table 5. Evidence synthesis of the content validity of PROMs: rating of results and overall level of evidence.

PROM	PROM development		Relevance*		Comprehensiveness*		Comprehensibility	
	PROM design	CI study	RoR	QoE	RoR	QoE	RoR	QoE
ADAP	D		+	low	?	very low	+	low
AEQ	A		+	very low	+	very low	+	very low
APS	V	I	+	very low	-	very low	+	low
BSFAQ				N/A				
CAS-D-65+	V		+	moderate	+	very low	±	very low
CPCI	D		+	low	+	very low	+	low
CPCI-42	D		+	low	±	very low	+	low
FABQ	V		?	very low	+	moderate	+	moderate
FACS	I	I	±	very low	+	very low	+	moderate
NRP	D		+	very low	±	very low	±	very low
PARQ	I		+	very low	+	very low	+	very low
PASS	D		+	very low	+	very low	+	very low
PASS-20	V		+	very low	+	very low	+	moderate
PCI	N/A		+	low	-	very low	+	very low
PIPS	V	I	±	very low	?	very low	?	very low
POAM-P	D		+	very low	±	very low	+	moderate
TSK	N/A		+	moderate	+	moderate	+	moderate
TSK-11	I		+	very low	±	very low	+	moderate

TSK-13				N/A				
TSK-TMD	D	D	+	low	±	very low	+	moderate

CI = Cognitive interview; QoE = Quality of evidence; RoR = Rating of results; * combined for patient and expert opinion

“V” = very good, “A” = adequate, “D” = doubtful, “I” = inadequate; “+” = sufficient, “–” = insufficient, “±” inconsistent, “?” = indeterminate.

Table 6. Risk of Bias Assessments and Rating Against Measurement Properties of included studies.

Instrument	Reference	structural validity	internal consistency	reliability	measurement error	criterion validity	convergent validity†	responsiveness to change‡
ADAP	Ansanello 2022 [8]	? / A	+ / V					
	Ansanello 2023 [9]			+ / D	? / D		a. - / V	
AEQ	Hasenbring 2009 [68]	? / I	? / V				a. + / A	
							b. + / D	
	Abad 2020 [2]		+ / V	+ / D	? / D		a. - / D	
	Ruiz-Párraga 2015 [170]	? / I	? / V				a. + / D	
APS	Esteve 2016 - study 1 [51]	+ / V	- / V				a. + / A	
	Esteve 2016 - study 2 [51]		+ / V				a. + / A	
BAT-Back	Holzapfel 2016 [72]		+ / V	+ / V			a. + / A	
							b. + / V	
	Küçükakkaş 2020 [100]	? / A	+ / V	+ / A			a. - / I	
							b. + / V	
BSFAQ	Chen 2023 [24]						a. + / A	
							b. + / I	
CAS-D-65+	Quint 2011 [157]		? / V	? / D			a. + / A	
clinician-reported FAB	Calley 2010 [21]			- / V			a. - / V	

CPCI	Jensen 1995 [78]		? / V	+ / D		a. + / A	
	Hadjistavropoulos 1999 [66]	? / I				a. + / V	
	Tan 2005 [184]	? / I				a. + / A	
	Souza 2021 [180]	+ / V	+ / V				
	Truchon 2006 [191]	+ / D	+ / V				
	Romano 2003 [166]		? / V	+ / D		a. + / A	d. + / I
	Souza 2018 [179]		? / V				
CPCI-42	Ko 2010 [88]	? / I	? / V	+ / D		a. - / I	
	Monticone 2013 [130]	+ / A	+ / V	+ / A		a. + / I	
	Wong 2010a [211]	? / A	? / V			a. + / A	
	Romano 2003 [166]		+ / V	+ / D	+ / V	a. + / A	d. + / I
	Misterska 2014 [123]		? / V	+ / D		a. + / A	
	Garcia-Campayo 2007 [55]	+ / V	+ / V	? / A		a. + / I	
FABQ	Georgoudis 2007 [60]	? / I	? / V	+ / D		a. - / A	d. - / A
	Pei 2010 [142]	- / V	? / V	+ / A		a. ? / I	d. - / I
	Staerkle 2004 [182]	- / V	? / V	+ / D		a. - / I	
	Lee 2006 [106]	? / A	? / I	+ / A		a. + / A	b. - / I
						b. + / D	d. - / D
	Panhale 2019 [140]	? / A	? / V	+ / D		a. ? / I	
	Matsudaira 2014 [111]	? / V	? / V			a. + / A	

					b. + / I	
Monticone 2012 [129]	? / A	? / V	+ / D	- / D	a. + / I	
Terho 2016 [185]	? / I	? / V	+ / I			
Ibrahim 2019 [73]	? / A	? / V	+ / D	? / D	a. - / I	d. + / V
Igwesi-Chidobe 2019 [74]	? / A	? / V	+ / D	- / D	a. - / I	
Chaory 2004 [23]	? / A		+ / D	? / D	a. + / I	d. + / D
Pfingsten 2000 [145]	? / A	? / V	? / D		a. + / I b. + / D	
Rostami 2014 [169]	? / A	? / V	+ / A		a. + / I	
Pfingsten 2004 [144]	? / A	? / V	? / D		a. + / I	
Mintken 2010 [122]	? / D	? / V	? / A		a. - / V	
Waddell 1993 [206]	? / A	+ / V	+ / D		a. + / I	
George 2010 [59]			+ / I	? / I	a. + / I	
Pruneti 2014 [154]	? / A	? / I	? / I			
Cleland 2008a [33]					a. + / A	
Cheung 2018 [25]		± / V			a. - / D b. - / D	
Riley 2019 [161]			- / D		a. - / V	b. ? / I d. + / I
Inrig 2012 [75]		+ / V	- / I	? / I	a. - / I	

Laufer 2012 [105]		+ / V	- / I			a. + / V b. + / A	
Alanazi 2017 [5]			+ / D			a. - / I	
Trolle 2019 [189]			+ / D	? / D		a. ? / V	b. + / A
Bid 2016 [18]		- / V	+ /D	? / D		a. - / I	
Bid 2019 [15]	- / V					a. - / I	d. + / I
Kaka 2015 [80]			+ / D		+ / V	a. - / I	
Kumar 2020 [99]		? / V	+ / D				
Panhale 2018 [141]		+ / V	+ / D			a. - / I	
Grotle 2006 [64]	? / I	? /V	± / I	? / I		a. + / I	d. + / D
Kovacs 2006 [91]		? / I	+ / D				
Dayalan 2022 [42]		+ / V	+ / V			a. + / I	
Dedering 2013 [45]		+ / I	+ / I			a. - / I b. - / V	
Franchignoni 2020 [52]	? / A	? / V					
Kamonseki 2021 [81]	? / A	? / V	- / A	? / A			d. - / D
Monticone 2020 [131]		? / V	+ / D	+ / D			c. + / I d. + / V
Namli 2022 [134]		? / V				a. ? / V	
Ozuberk 2023 [138]		? / V	+ / D			a. - / D	

	Pagels 2023 [139]		? /V			a. - / I b. - / D	
	Saadat 2023 [172]						c. + / A d. + / V
	Salama 2020 [173]					a. + / A	
	Sørensen 2021 [183]						b . + / D d. + / V
	Korkmaz 2009 [90]	? / V	+ / V	? / D		a. + / I	d. - / V
	Mbada 2021 [112]	? / V	+ / I	+ / D	? / D	a. - / I	
	Cleland 2008b [34]			+ / I		a. + / V	
	de Souza 2008 [44]		? / V	+ / D		a. + / I	b. - / I d. - / I
	Pfingsten 1997 [146]	? / D	? / I			a. - / I	
	Askary-Ashtiani 2014 [12]		? / V	+ / D		a. + / I b. + / A	
	Crombez 1999 [38]		? / V			a. - / I	
FACS	Neblett 2016 [135]		? / V	? / D		a. + / I b. + / A	
	Cuesta-Vargas 2020 [40]	? / I	? / V			a. + / V	
	Aroca 2023 [62]					a. - / V	

	De Baets 2023 [43]	? / A	? / V	+ / D	? / D		a. + / V	
	Bid 2020 [16]		? / V	+ / A	? / A		a. + / A	
	Duport 2023 [46]		? / V	+ / A			a. + / V	
	Turan 2023 [192]	? / A	? / V	+ / D			a. - / D	
	Knezevic 2018 [87]	+ / V	+ / V	+ / D	? / D		a. + / I b. + / V	
	Neblett 2017 [136]	? / A						d. - / D
NRP	Jensen 2017 [79]	+ / V	+ / V	? / D			a. + / A	
PARQ	McCracken 2007 [114]	? / A	? / V				a. + / A	
PASS	McCracken 1992 [116]		? / V				a. + / I	
	Crombez 1999 [38]		? / V				a. - / I	
PASS-20	Coons 2004 [35]	? / A	? / V	? / I			a. + / D	
	Kreddig 2015 [94]	? / A	? / V				a. + / A	
	Zhou 2017 [213]	- / V	? / V	+ / D			a. + / D	
	Cho 2010 [27]	- / V	? / V	? / I			a. + / A	
	McCracken 2002 [113]		? / V			+ / V	a. + / D	
PCI	Kraaimaat 2003 [93]		? / V	? / I			a. + / I b. + / V	
PIPS	Vasiliou 2019 [197]	- / V	? / V	? / I			a. + / I	
	Wicksell 2010 [208]	+ / V	+ / V				a. + / I	

	Barke 2015 [13]	+ / V	+ / V			a. + / I	
	Trompetter 2014 184]	+ / V	+ / V			a. + / I	
	Nagasawa 2021 [133]	+ / V	+ / V	? / D		a. ? / D	
	Rodero 2013 [162]	? / A	? / V	+ / D		a. + / A	
	Wicksell 2008 [209]	? / A	+ / V			a. + / A	
POAM-P	Cane 2013 [22]		? / V			a. + / A	d. ? / I
	Benaim 2017 [14]		? / V	+ / D		a. + / I	
TSK	Lundberg 2004 [110]	? / A	? / I	+ / A		b. + / V	
	Goubert 2004 [63]	? / V					
	Bid 2018 [17]	? / A	? / I	- / D	? / D	a. - / A	
	Monticone 2010 [132]	? / A	? / V	+ / D		a. + / I	
	French 2007 [53]	- / V	? / V			a. + / D	d. ? / I
	Roelofs 2004 [163]	- / V	? / V			a. + / I	
	Wei 2015 [207]		? / V	+ / D		a. + / A	
	Roelofs 2007 [164]	+ / V					
	Areeudomwong 2017 [10]		? /V	+ / D		a. - / I	
	Alanazi 2021 [6]		? / V	+ / A		a. + / D	
	Dedering 2013 [45]		? / V	+ / I		a. - / I b. + / V	
	Koho 2014 [89]		? / V	+ / D			

Monticone 2016a [127]				+ / V		b. + / I c. + / D
Monticone 2016b [128]				+ / D		b. + / I c. + / D d. + / V
Pulles 2020 [155]				- / D		c. - / D d. - / V
Cleland 2008b [34]			+ / I		a. + / I	
de Souza 2008 [44]		? / V	+ / D		a. + / I	b. + / I d. + / I
Wong 2010b [212]	- / V	- / V			a. + / D b. + / V	
Lamé 2008 [102]		+ / V	- / I			
Woby 2005 [210]		? / I	+ / I	? / I		b. + / V c. + / D d. + / V
Kikuchi 2015 [86]		+ / I			a. + / I b. + / D	
Vlaeyen 1995a [199]		+ / V			a. + / I	
Heuts 2004 [71]	- / V					

	Siqueira 2007 [177]	? / D	? / V					
	Georgoudis 2022 [61]		? / V	+ / D			a. + / D	
	Eiger 2023 [48]			+ / V	? / V			
	Vlaeyen 1995b [200]	? / A	- / V				a. + / I b. + / D	
	Crombez 1999 [38]		? / V				a. + / I	
	Burwinkle 2015 [19]	? / A	? / V					
	Askary-Ashtiani 2014 [12]		? / V	+ / D			a. + / I b. + / A	
	Saadat 2023 [172]							c. + / A d. + / V
	Haugen 2008 [69]	? / A	? / V	- / I			a. + / I	c. ? / D
TSK-11	Rusu 2014 [171]	+ / V	- / V					
	Gómez-Perez 2011 [65]	? / A	? / V	? / I			a. - / A b. - / V	
	Tkachuk 2012 [188]	- / V	? / V				a. + / A	
	Larsson 2014 [104]	- / V	? / V	+ / A			a. + / I	
	Mintken 2010 [122]	? / D	? / V	+ / A			a. - / V	
	Woby 2005 [210]		? / I	+ / I	? / I			a. + / V b. + / V

								c. + / D
								d. + / V
	Cai 2019 [20]	? / A	? / V	+ / D			a. - / A	
	Kamonseki 2021 [81]	? /A	? / V	+ / A	? / A			d. - / D
	Eiger 2023 [48]			+ / V	? / V			
	Santo Salvador 2021 [174]		? / V	+ / D	+ / D	+ / D	a. - / V	
	Kikuchi 2015 [86]		? / I				a. + / I	
							b. + / D	
	Satpute 2019 [175]		+ / H	+ / D			a. + / D	
	Roelofs 2007 [164]	+ / V	- / V				a. + / A	
	Hapidou 2012 [67]			+ / D	? / D		a. + / V	c. + / I
	George 2010 [59]			+ / D	? / D		a. - / I	
	Geisser 2000 [56]	? / V					a. - / I	
	Tkachuk 2012 [188]	- / V						
	Cordeiro 2013 [36]	? / I	+ / I	+ / D			a. + / I	d. - / D
TSK-13	Goubert 2004 [63]	? / V	+ / V					
	Heuts 2004 [71]	+ / V	+ / V					
	Burwinkle 2015 [19]	? / A	? / V					
	Eiger 2023 [48]			+ / V	? / V			
	Wei 2015 [207]	? / A	? / I	+ / D			a. + / A	

TSK-TMD	LaTouche 2020 [101]	+ / V	+ / V	+ / D	? / D	a. + / A
	Aguiar 2017 [4]	+ / V	+ / V	+ / D	? / D	a. + / D
	Visscher 2010 [198]	+ / V	+ / V	- / A		a. ? / A
	He 2016 [70]		? / V	+ / D		a. ? / I

“V” = very good, “A” = adequate, “D” = doubtful, “I” = inadequate; “+” = sufficient, “–” = insufficient, “±” inconsistent, “?” = indeterminate.

† a. convergent validity; b. known-groups validity.

‡ a. comparison with gold standard; b. comparison with other instruments; c. comparison between subgroups; d. comparison before and after intervention.

Table 7. Evidence synthesis of the validity of PROMs: rating of results, overall level of evidence and recommendation for use.

Rec.	PROM	Language	Validity			Reliability			Responsiveness
			Structural	Construct	Criterion	Internal consistency	Reliability	Measurement Error	
A	CPCI	Brazilian	+ (H)			+ (M)			
	CPCI-42	Italian	+ (M)	+ (VL)		+ (M)	+ (M)		
B	ADAP	Brazilian	? (M)	- (L)		? (H)	+ (L)	? (L)	
	AEQ	German	? (VL)	+ (M)		? (H)			
		Iranian		- (L)		? (H)	+ (L)	? (L)	
		Spanish	? (VL)	+ (M)		? (H)			
	APS	Spanish	+ (H)	+ (H)		± (M)			
	BAT-Back	German		+ (H)		? (H)	+ (H)		
		Turkish	? (M)	± (VL)		? (H)	+ (M)		
	BSFAQ	English / French		+ (VL)					
	CAS-D-65	German		+ (L)		? (H)	? (VL)		
	clin.-rep. FAB	English		- (H)			- (H)		
	CPCI	English	? (VL)	+ (H)		? (H)	+ (L)		+ (VL)
		French	+ (H)			+ (VL)			
	CPCI-42	English	+ (H)	+ (M)	+ (H)	? (H)	+ (M)		+ (VL)
		Chinese	? (M)	+ (M)		? (M)			
		Korean	? (VL)	- (L)		? (VL)	+ (L)		
		Polish		+ (M)		? (H)	+ (L)		
	FABQ	English	? (M)	? (M)		? (M)	+ (M)	? (L)	+ (M)
		Arab		? (M)		? (H)	? (L)		
		Brazilian	? (M)	+ (VL)		? (H)	+ (M)	? (L)	- (L)
		Danish							+ (L)
		Dutch		- (VL)		? (H)			
		Finnish	? (VL)			? (VL)	+ (VL)		
		French	? (M)	+ (VL)			+ (L)	? (L)	+ (L)
		German	? (H)	± (L)		? (H)	? (M)		
		Greek	? (VL)	- (M)		? (VL)	+ (L)		- (M)
		Hausa	? (M)	- (L)	+ (H)	? (M)	+ (M)	? (L)	+ (H)
		Hindi	? (M)	? (VL)		? (M)	+ (L)		

	Igbo	? (M)	- (VL)		? (M)	+ (L)	- (L)	
	Italian	? (H)	+ (VL)		? (H)	+ (M)	? (M)	+ (M)
	Japanese	? (H)	+ (M)		? (H)			
	Kannada				? (H)	+ (L)		
	Marathi		- (VL)		? (H)	+ (L)		
	Norwegian	? (VL)	+ (VL)		? (H)	± (VL)	? (VL)	+ (L)
	Persian	? (M)	+ (M)		? (M)	+ (M)		+ (M)
	Spanish	? (VL)				+ (L)		
	Swedish		- (VL)		? (VL)	+ (VL)		
	Tamil		+ (VL)		? (M)	+ (M)		
	Turkish	? (H)	± (M)		? (H)	+ (L)		- (H)
	Yoruba	? (H)	- (VL)		? (VL)	+ (L)	? (L)	
FACS	Dutch	? (M)	+ (H)		? (H)	+ (L)	? (L)	
	English	? (M)	+ (M)		? (H)	? (L)		- (L)
	French		+ (M)		? (H)	+ (L)		
	Gujarati		+ (M)		? (H)	+ (M)	? (M)	
	Spanish	? (VL)	± (M)		? (VL)			
	Serbian	+ (H)	+ (VL)		+ (H)	+ (L)		
	Turkish	? (M)	- (L)		? (H)	+ (L)		
NRP	English	+ (H)	+ (M)		+ (H)	? (L)		
PARQ	English	? (M)	+ (M)		? (H)			
PASS	English		+ (L)		? (H)			
	Dutch		- (VL)		? (H)			
PASS-20	English	? (M)	+ (M)	+ (H)	? (H)	? (VL)		
	German	? (M)	+ (M)		? (H)			
	Korean	- (VL)	+ (L)		? (H)	+ (VL)		
PCI	Dutch		+ (VL)		? (L)	? (VL)		
PIPS	Dutch	+ (H)	+ (VL)		+ (H)			
	German	+ (H)	+ (L)		+ (H)			
	Japanese	+ (H)	? (L)		+ (H)	? (L)		
	Spanish	? (M)	+ (L)		? (H)	+ (L)		
	Swedish	+ (H)	+ (M)		+ (H)			
POAM-P	English		+ (L)		? (H)			? (VL)
	French		+ (VL)		? (H)	+ (L)		
TSK	English	± (H)	+ (H)		? (H)	+ (M)	? (VL)	+ (H)

		Arabic		+ (VL)		? (M)	+ (L)		
		Brazilian		+ (VL)		? (H)	+ (L)		+ (VL)
		Danish					+ (H)	? (H)	
		Finnish				? (M)	+ (VL)		
		Greek		+ (L)		? (H)	+ (L)		
		Gujarati	? (M)	- (M)		? (VL)	- (L)	? (L)	
		Japanese		+ (M)		+ (VL)			
		Norwegian	? (M)	+ (VL)		? (H)	- (VL)		? (L)
		Persian		+ (M)		? (H)	+ (L)		+ (M)
		Swedish		- (VL)		? (L)	+ (VL)		
		Thai		- (VL)		? (M)	+ (VL)		
	TSK-11	Brazilian	? (M)	- (H)	+ (H)	? (H)	+ (M)	± (L)	- (VL)
		Chinese	? (M)	- (M)		? (H)	+ (L)		
		Danish					+ (H)	? (H)	
		Japanese		+ (M)		? (VL)			
		Marathi		+ (L)		+ (H)	+ (L)		
		Spanish	? (M)	- (M)		? (M)	? (VL)		? (VL)
	TSK-13	Danish					+ (H)	? (H)	
		Dutch	+ (M)			+ (H)			
		Portuguese	? (VL)	+ (VL)		? (VL)	+ (L)		- (L)
	TSK-TMD	Dutch	+ (H)	? (M)		+ (H)	± (M)		
		Brazilian	+ (H)	+ (L)		+ (H)	+ (L)	? (L)	
		Chinese		? (VL)		? (H)	+ (L)		
		Spanish	+ (H)	+ (L)		+ (H)	+ (L)	? (L)	
C	FABQ	Chinese	- (H)	+ (L)		? (H)	+ (H)		- (L)
		Gujarati	- (H)	- (H)		? (H)	+ (L)	? (L)	+ (VL)
		Swiss-German	- (H)	- (VL)		? (H)	+ (L)		
	PASS-20	Chinese	- (H)	+ (L)		? (H)	+ (L)		
	PIPS	Greek	- (H)	+ (VL)		? (H)	? (VL)		
	TSK	Chinese	- (H)	+ (H)		? (M)	+ (L)		
		Dutch	- (H)	+ (M)		- (H)		- (H)	- (M)
		Italian	? (M)	+ (VL)		- (H)	+ (L)	+ (H)	+ (H)
	TSK-11	English	- (H)	± (M)		? (H)	+ (H)	? (L)	+ (H)
		Dutch	+ (H)	+ (M)		- (H)			
		German	+ (H)			- (H)			

	Swedish	- (H)	+ (L)		? (H)	+ (M)
TSK-13	English	- (H)	- (VL)		? (H)	

Rec. = Recommendation; Responsiveness = Responsiveness to change; A = PROM can be recommended for use; B = PROM has potential to be recommended for use; C = PROM should not be recommended for use; “H” = high level of evidence, “M” = moderate level of evidence, “L” = low level of evidence, “VL” = very low level of evidence; “+” = sufficient, “−” = insufficient, “±” inconsistent, “?” = indeterminate.