Journal of Hand Therapy 32 (2019) 222–232

Contents lists available at ScienceDirect

Journal of Hand Therapy

journal homepage: www.jhandtherapy.org

JHT Read for Credit Article #610.
Scientific/Clinical Article

Establishing the psychometric properties of 2 self-reported outcome measures of elbow pain and function: A systematic review

Joshua I. Vincent PT, PhD [a,b,c,*], Joy C. MacDermid PT, PhD [d,e], Graham J.W. King MD, FRCSC [f,g], Ruby Grewal MD, FRCSC [h,i], Emily Lalone PhD [j]

[a] School of Rehabilitation Sciences, McMaster University, Hamilton, Ontario, Canada
[b] Roth-McFarlane Hand and Upper Limb Centre, St. Joseph's Hospital, London, Ontario, Canada
[c] Lifemark Health Corp., Toronto, Ontario, Canada
[d] University of Western Ontario, School of Physical Therapy, London, Ontario, Canada
[e] Roth-McFarlane Hand and Upper Limb Centre, St. Joseph's Hospital, London, Ontario, Canada
[f] Department of Surgery, University of Western Ontario, London, Ontario, Canada
[g] Roth-McFarlane Hand and Upper Limb Center, St. Joseph's Hospital, London, Ontario, Canada
[h] Department of Surgery, University of Western Ontario, London, Ontario, Canada
[i] Roth-McFarlane Hand and Upper Limb Center, St. Joseph's Hospital, London, Ontario, Canada
[j] Faculty of Engineering, University of Western Ontario, London, Ontario, Canada

ARTICLE INFO

Article history:
Received 27 March 2018
Received in revised form
15 July 2018
Accepted 20 July 2018
Available online 23 December 2018

Keywords:
PREE
pASES-e
Clinical measurement properties
Systematic review

ABSTRACT

Study Design: Systematic review.
Introduction: The Patient-Rated Elbow Evaluation (PREE) and the self-report section of the American Shoulder and Elbow Surgeons—elbow form (pASES-e) are 2 patient-reported outcome measures (PROMs) commonly used to assess pain and disability arising from elbow disorders.
Purpose of the Study: To systematically review and summarize the quality and content of the evidence that is available on the psychometric properties of the PREE and pASES-e.
Methods: We systematically searched the online databases PubMed, EMBASE, ProQuest, Scopus, Cumulative Index to Nursing and Allied Health Literature, UptoDate, ProQuest Dissertations & Theses, and Google Scholar. Ninety-one articles were retrieved, and after screening, 9 were included in the final analysis. Data extraction and quality appraisal was performed by 2 independent raters. Descriptive synthesis of the reviewed studies was completed.
Results: Seven of the 9 studies had a quality score of 75% or higher. Agreement between the raters was good (kappa, 0.81). Both the PROMs did not demonstrate any floor and ceiling effects except for the satisfaction subscale of the pASES-e. Factor analysis revealed multidimensionality in the function subscale for both the PROMs. Construct validity was good with correlations above 0.70. Both were highly reliable with interclass correlation coefficient of >0.90. They were also highly responsive with an effect size and standardized response mean above 1. The minimal clinical important difference was not estimated for either measures.
Discussion: This study concluded that strong clinical measurement properties exist for both the PREE and the pASES-e. We identified gaps in the current evidence for both the ASES-e and the PREE. Future studies need to calculate clinically important estimates like MCID, SEM, and others; and provide clear and specific conclusions.
Conclusion: The PREE and pASES-e have been established to be valid, reliable, and sensitive to change in both clinical and research settings based on high-quality evidence.

Introduction

Patient-reported outcome measures (PROMs) are commonly used to measure pain and disability. This is consistent with the policy of patient-centered care where patients are empowered to make

No funds were received in support of this study.
* Corresponding author. Roth-McFarlane Hand and Upper Limb Centre, St. Joseph's Healthcare London, 268 Grosvenor Street, London, Ontario, Canada N6A 4L6. Tel.: +1 519 694 6222; fax: +1 519 646 6049.
E-mail address: joshuaisrael1985@gmail.com (J.I. Vincent).

0894-1130/$ – see front matter © 2018 Hanley & Belfus, an imprint of Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.jht.2018.07.004

informed decisions regarding their health status and its management.[1,2] PROMs should be able to accurately assess a person's health status and detect change in the state of health as it evolves. Hence, these measures require rigorous evaluation of their measurement properties. It is important for the clinicians to ensure that the PROMs they use are valid and reliable, so that they can be confident about the results they obtain and the inferences they make.[3] Clinical utility is another issue to consider with the use of PROMs. Administrative burden (ie, time, cost, and ease of calculating final scores) and patient burden (ie, reading level, ease of completion, time required) are important factors that determine the clinical utility of a PROM.

Evidence about the clinical measurement properties of a PROM requires testing of different clinical measurement properties, patient populations, and contexts; hence it is rarely accomplished in a single study. Therefore, it is important to identify, evaluate, and synthesize the pool of evidence that establishes a greater understanding of measurement performance of different tools. Systematic reviews of clinical measurement studies should provide an unbiased synthesis of the body of literature supporting specific applications of different PROMs and thereby assist clinicians by accessing the best evidence for selection and application of PROMs for their practice.

Two joint-specific PROMs that are commonly used to assess pain and disability that follow elbow disorders are the Patient-Rated Elbow Evaluation (PREE)[4,5] and the self-report section of the American Shoulder and Elbow Surgeons—elbow form (pASES-e).[6] Recently, the PREE has been recommended as one of the measures in the score set to assess patients with elbow disorders.[7] This score set was established for the purpose of measuring quality of life and objective and subjective functions in general elbow pathologies.[7] Cross-cultural adaptations of these tools are also emerging.[8-10] For the aforementioned reasons, we have included the PREE and pASES-e in the current systematic review.

A previous systematic review has focused on evaluating several rating systems of the elbow.[11] There were some potential pitfalls in that study. First, they had a broad research question with a strict inclusion and exclusion criteria, which may have resulted in exclusion of studies that could have potentially driven the conclusion of the review.[11] Furthermore, the authors of this review only searched 1 database (PubMed).[11] This created the potential for selection bias and missing evidence that might have been obtained through other electronic databases (ie, Scopus, ProQuest). Previous studies have shown that a search of PubMed alone does not identify all relevant articles.[12,13] The et al[11] did not include studies from languages other than English or non-English versions of the PROMs. This may have led to language bias. Some of the articles that were missed in this systematic review had the potential to change the conclusions on the quality of the PREE. The pASES-e was not included in the review. Hence, the purpose of the present study was to perform a comprehensive systematic review of the literature and summarize the quality and content of the evidence that is available on the psychometric properties of the PREE and pASES-e.

Materials and methods

Description of the measures

Patient-rated elbow evaluation

The PREE was developed in the year 2000. It is a 20-item PROM that evaluates pain and disability arising from elbow disorders.[4,5] It has 2 sections, namely, pain and function. The pain section (5 items) is scored on an 11-point ordinal scale where 0 represents *did not have any pain* and 10 represents *the worst pain you have ever experienced*. The function section has 15 items. The function section is further subdivided into specific activities (11 items) and usual activities subscales (4 items). Each of the items is scored on an 11-point ordinal scale where 0 represents *no difficulty* and 10 represents *unable to do*. The total score is calculated out of 100, where pain and disability are equally weighted. The higher the PREE total score, the greater the pain and disability.[14]

American shoulder and elbow surgeons—elbow form (self-report section)

The pASES-e is an 18-item PROM with 3 sections—pain, function, and satisfaction.[6] The pain section (5 items) is scored on an 11-point ordinal scale, where 0 is no pain and 10 is the worst pain ever. The function section (12 items) evaluates both the affected and unaffected arms and is scored on a 4-point ordinal scale with 0 (unable to do) and 3 (not difficult). The greater the total score, the greater the disability. The satisfaction section has 1 item asking patients to rate their satisfaction from treatment on an 11-point ordinal scale where 0 represents *not at all satisfied* and 10 *very satisfied*. There is no total score for the pASES-e.

Search strategy

Literature search was completed by one of the authors (JIV); search yield was verified with another author (JCM). Electronic databases PubMed, EMBASE, Scopus, ProQuest, Cumulative Index to Nursing and Allied Health Literature, UptoDate, ProQuest Dissertations & Theses, and Google Scholar along with specific journals like *Hand*, *Journal of Orthopaedic & Sports Physical Therapy*, and the *Journal of Hand Therapy* were searched; also reference lists were retrieved from articles and used to identify potential articles that could have been missed during the regular search. Because these questionnaires were developed after 1995, these databases were searched from January 1995 to May 2015, in all languages. Three groups of search terms were used in various combinations using ANDs and ORs (Appendix A). The Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines[15] were followed in reporting the search strategy (Fig. 1).

Selection criteria

Articles were included in the review if they met the following inclusion criteria: (1) measurement of at least 1 psychometric property of either the PREE or the pASES-e (clinical measurement properties are listed in Table 1) and (2) inclusion of subjects with any elbow disorders. Studies were excluded if there were no statistical values mentioned to verify the results. The title/abstracts from the search yield were screened by 2 independent reviewers based on the inclusion criteria. Full articles were retrieved for the selected articles. After this, the final set of articles that met the inclusion criteria were selected by the 2 reviewers and subjected to quality appraisal. There was no discrepancy among the reviewers in study selection.

Quality appraisal

A pair of raters independently conducted a quality appraisal of the articles retrieved based on the inclusion criteria. The quality appraisal of the included studies was assessed using a previously developed quality appraisal tool developed by MacDermid—the Quality Appraisal for Clinical Measurement Studies (QA-CMS) (Appendix A).[16] The QA-CMS has demonstrated high inter-rater reliability (interclass correlation coefficient [ICC], >0.90).[17,18] Once the quality was appraised, data were extracted using a data extraction form.[16] The reviewers met with the developer of the quality appraisal tool (JCM) and performed a calibration review. Every item of the appraisal tool was briefed to the reviewers to clarify the intended meaning and interpretation. Then the articles were independently appraised by the 2 reviewers. Once this process was complete, both the reviewers met for discussion to achieve consensus on individual items of appraisal. When consensus was not achieved, a third reviewer helped to resolve the disagreement. The studies were arranged in the descending order of quality scores. The

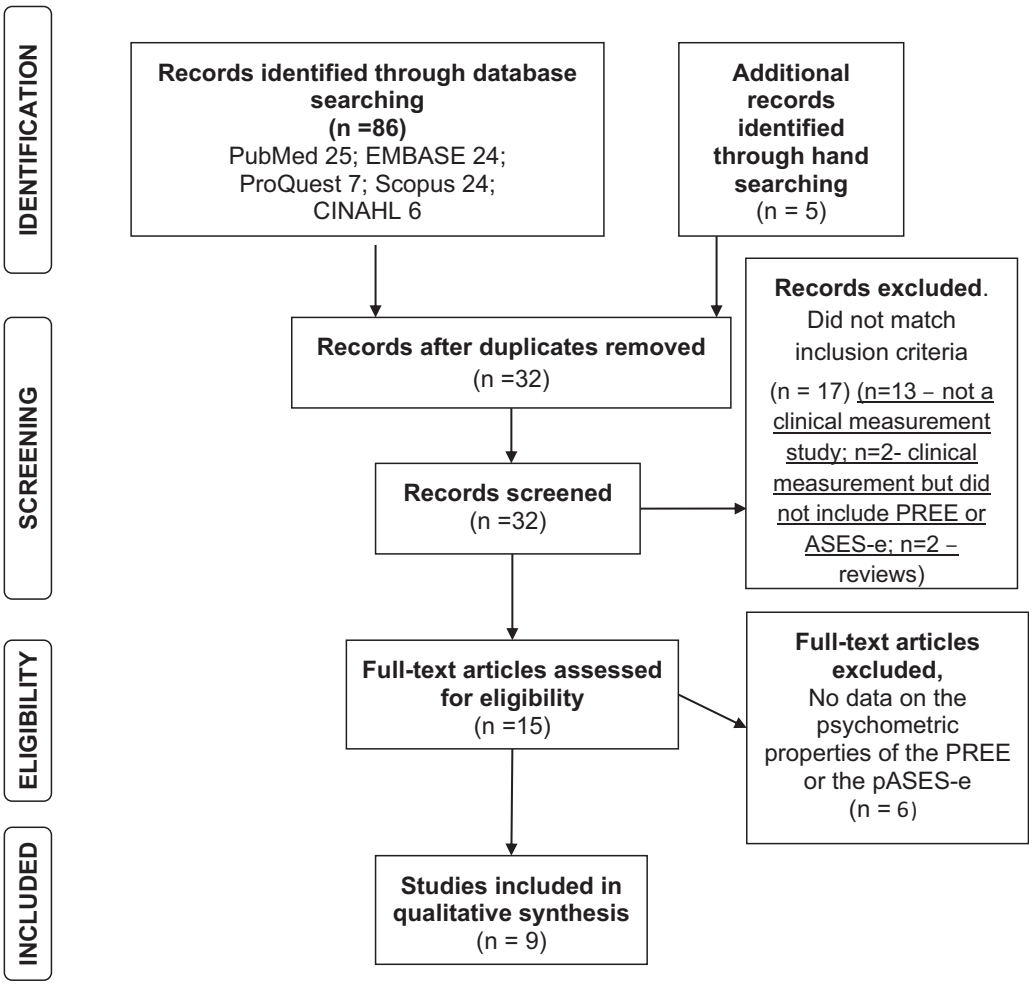

Fig. 1. Systematic review evidence flowchart based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. CINAHL = Cumulative Index to Nursing and Allied Health Literature; PREE = Patient-Rated Elbow Evaluation; pASES-e = self-report section of the American Shoulder and Elbow Surgeons—elbow form.

kappa value was calculated to assess the overall inter-rater agreement before consensus was achieved.[19,20] The inter-rater reliability of the quality assessment tool was measured using ICC.[21-24]

Results

Search yield

The search of the aforementioned sources yielded 91 articles. After the removal of duplicates and irrelevant articles, 32 articles remained for screening using the inclusion and exclusion criteria. Seventeen of the 32 articles did not meet the selection criteria. Full text was retrieved for 15 articles. After reviewing the full text, 6 additional articles were excluded. Finally, 9 articles were then included for quality appraisal (Fig. 1). The PREE was studied more frequently than the pASES-e. Of these 9 articles, 4 focused specifically on the PREE[7,8,10,25]; only 1 was specific to pASES-e[9]; and the remaining 4 included both the PROMs.[5,26-28] The synthesis of individual studies and psychometric properties measured are described in Table 2. All studies included patients who underwent surgical intervention.

Quality rating

The quality score for the included articles ranged from 54%[27] to 96%,[7] with 78% (7 of 9) of articles having a score of 75% or higher (Table 3). The quality assessment tool exhibited excellent inter-rater reliability (ICC, 0.94; 95% confidence interval [95% CI], 0.91-0.96). The overall agreement between the raters was good (kappa, 0.81; $P < .001$). The common design flaws observed with the reviewed articles were that none of the studies completed a proper sample size calculation. Only 5 of the 9 studies explicitly reported their hypotheses.[7-9,16,26] The studies also lack clarity in reporting the type and subtype of the psychometric properties being measured. For example, most of the studies mentioned that they were assessing validity and reliability in their objective but did not explicitly mention the type of validity or reliability they were intending to measure except for 2 studies.[16,26] In addition, most studies did not report the estimates of measurement error like the standard error of measurement (SEM) or use CIs. Except for the study by John et al,[9] none of the studies reported SEM. As well, conclusions were vague and nonspecific.

Readability, language, and cross-cultural translation/adaptation

Both the PREE and pASES-e have been reported to be easy to read and understand (Table 4). The reading level for any outcome measure is measured based on the Flesch-Kincaid grade level of readability.[29] PREE was equivalent to that of grade 6 (Flesch-Kincaid grade level of readability, 6.3).[30] With

Table 1
Definitions of clinical measurement properties reviewed in this systematic review

Clinical measurement property	Definition
Test-retest reliability	The degree to which an instrument is stable based on repeated administrations of the test to the same individual during a specified time interval
SEM	A reliability measure of response stability, estimating the standard error in a set of repeated scores
MDC	It is the statistical estimate of the smallest amount of change that can be detected by an outcome measure
Internal consistency	A form of reliability; assessing the degree to which a set of items in an instrument all measure the same trait
Construct validity	A type of measurement validity; the degree to which a theoretical construct is measured by an instrument
Convergent validity	An approach of construct validation, assessing the degree to which 2 different instruments or methods are able to measure the same construct
Discriminant or divergent validity	An approach in construct validation assessing the degree to which an instrument yields different results when measuring 2 different constructs; that is, the ability to differentiate between the constructs
Known group validity	A technique for construct validation, in which validity is determined by the degree to which an instrument can demonstrate different scores for groups known to vary on the variable being measured
Responsiveness	The ability of a test to demonstrate change
Factor validity	An exploratory multivariate statistical technique used to examine the structure within a large set of variables and to determine the underlying dimensions that exist within that set of variables
Floor and ceiling effects	A ceiling effect occurs when a high proportion of subjects in a study have maximum scores on the observed variable. A floor effect occurs when a large concentration of participant scores are at the lower limit of the scale

SEM = standard error of measurement; MDC = minimal detectable change.

regard to missing responses to items, a minimal completion criterion of 65% was set for the PREE.[10] There was no specific pattern of missing responses to items.[8,10] For the pASES-e, 1 study reported that items 18 and 19 (do usual work or sport) (27.5%-33%) were missing.[9] The PREE has been translated into German[10] and Japanese[8] and has been reported to have exhibited psychometric properties that were equivalent to the English version. The pASES-e has also been translated into German exhibiting psychometric properties equivalent to that of the original English version.[9]

Administrative burden

Administrative burden was reported in only 1 study. Angst et al[28] reported that it took only 3 minutes each to complete the PREE and pASES-e.

Floor and ceiling effects

The floor and ceiling effects of the 2 questionnaires have been reported. Angst et al[28] observed a ceiling of 15%, 4%, and 4% for

Table 2
Summary table describing the studies included in the systematic review

Study	Quality score	PRO	Population	*n*	Clinical measurement properties measured
Angst et al[27]	95.83	ASES-e, PREE	Total elbow arthroplasty (RA and post-trauma cases) M:F = 19:46 Age; mean (SD) = 61.9 (13.0)	65	Responsiveness (SRM and ES), ROC curve analysis, test-retest reliability, sensitivity and specificity
Vincent et al[26]	87.5	PREE, ASES-e	Various elbow surgeries M:F = 84:44 Age; mean (SD) = 46.5 (12.8)	128	Internal consistency, construct validity, longitudinal validity, responsiveness (SRM and ES), factor analysis
Hanyu et al[8]	83.33	PREE	Various elbow pathologies M:F = 39:35 Age; mean (SD) = 46.7 (20.7)	74	Translation and cross-cultural adaptation, test-retest reliability, internal consistency, construct validity, responsiveness, factor analysis
MacDermid et al[5]	79.17	PREE, ASES-e	Elbow trauma (operative and nonoperative cases); infection; OA; RA; lateral epicondylitis M:F = 33:37 Age; mean (SD) = 49 (16)	70	Test-retest reliability, construct validity, divergent validity
John et al[9]	79.17	ASES-e	Total elbow arthroplasty (RA and post-trauma cases) M:F = 22:53 Age; mean (SD) = 64.1 (13.3)	75	Translation and cross-cultural adaptation, test-retest reliability, internal consistency, construct validity, divergent validity, floor and ceiling
Angst et al[28]	75	PREE, ASES-e	Total elbow arthroplasty (RA and post-trauma cases) M:F = 23:56 Age; mean (SD) = 64.1 (13.3)	79	Construct validity, exploratory factor analysis, known group validity
John et al[10]	75	PREE	Elbow prosthesis M:F = 19:37 Age; mean (SD) = 63.7 (11.4)	56	Translation and cross-cultural adaptation, test-retest reliability, internal consistency, construct validity, divergent validity
Schimdt and Stangl[25]	70.83	PREE	Various elbow pathologies M:F = 39:16 Age; mean (range) = 46 (13-71)	55	Construct validity
Liem et al[7]	54.16	PREE	Various elbow pathologies M:F = 33:33 Age; mean (range) = 49.99 (19.4-72.5)	66	Construct validity

PRO = patient-reported outcome; ASES-e = American Shoulder and Elbow Surgeons—elbow form; PREE = Patient-Rated Elbow Evaluation; RA = rheumatoid arthritis; M = male; F = female; SD = standard deviation; SRM = standardized response mean; ES = effect size; ROC = receiver operating characteristic; OA = osteoarthritis.

Table 3
Quality of studies included in the systematic review (arranged according to study quality)

Study	Item evaluation score for each criteria on the MacDermid quality assessment tool (minimum = 0; maximum = 2)												Total score	Quality score (%)
	1	2	3	4	5	6	7	8	9	10	11	12		
Angst et al[27]	2	2	2	2	2	1	2	2	2	2	2	2	23	95.83
Vincent et al[26]	2	2	2	2	1	2	2	1	2	2	2	1	21	87.50
Hanyu et al[8]	2	2	2	2	0	1	2	2	2	2	2	1	20	83.33
MacDermid et al[5]	2	1	2	2	0	1	2	2	2	2	2	1	19	79.17
John et al[9]	2	1	2	2	0	1	2	2	2	2	2	1	19	79.17
Angst et al[28]	2	1	2	2	0	N/A	2	2	2	2	1	2	18	75.00
John et al[10]	2	1	1	2	0	1	2	2	2	2	2	1	18	75.00
Schimdt & Stangl[25]	1	2	2	1	0	N/A	2	2	2	2	2	1	17	70.83
Liem et al[7]	1	2	2	0	0	N/A	1	1	2	2	1	1	13	54.16

N/A = not applicable to the study.
Disagreement in 30 instances: 78/108 ×; 100 = 72.22 before consensus.

the pain subscale, function subscale, and total PREE score, respectively, whereas observing a ceiling of 14% for the pain subscale, 4% for function-affected subscale, and 43% for satisfaction subscale. For the Japanese version of the PREE, no or very low (4%) floor and ceiling effects were observed,[8] whereas for the German version of the PREE, the ceiling ranged from 5.3% to 14.3%.[10]

Factorial validity

The PREE pain scale has been reported to be unidimensional both in the English version[26] and the Japanese version.[8] But the function subscale of the English version loaded onto 3 factors, whereas its Japanese counterpart loaded onto 2 factors.[8,26] For the pASES-e, the principle component analysis with varimax rotation revealed that the pASES-e loaded onto 4 different factors.[26]

Construct validity

Construct convergent validity has been established for the PREE and pASES-e. Their strength of relationship with the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire, QuickDASH, and the 36-Item Short Form Health Survey (SF-36) was consistently studied. For the PREE, high correlations ($r > 0.70$) were reported with the QuickDASH, DASH, pASES-e pain subscale, pASES-e total score, PREE pain subscale, and PREE function subscale.[5,7,8,10,25,26,28] Moderate correlations ($r = 0.30$-0.70) have been established with the clinician section of the ASES-e, pASES-e function subscale, and SF-36 physical component subscale.[5,10,26,28] The pASES-e total score correlated highly with the DASH, PREE function score, PREE total score,[9] and moderately with the clinician part of the ASES and SF-36 physical component subscale.[5,9,28] Construct divergent validity for both the PROMs has been established with the mental component summary score of the SF-36.[5,7-10,26,28] The strength of correlations of individual subscales with various constructs is shown in Table 4.

Reliability

Internal consistency

The internal consistency of the PREE has been shown to be slightly higher with the pASES-e. For the PREE, Vincent et al[26] reported a Cronbach alpha of 0.95 (95% CI, 0.94-0.96), whereas Hanyu et al[8] and John et al[10] reported 0.97 and 0.96, respectively, for the Japanese and German versions. For the pASES-e, the Cronbach alpha value was reported to be 0.93 (95% CI, 0.91-0.94)[26] for the English version and 0.90 for the German version[9] (Table 5).

Test-retest reliability

Relative reliability of the PREE has been well established and shown to have excellent reliability coefficients.[5,8,10] ICCs of 0.95 (95% CI, 0.86-0.98) for the total score and 0.88-0.89 for the individual subscales have been reported for the English version of the PREE.[5] For the Japanese and German versions, the ICCs varied from 0.73 to 0.94[8,10] for the different subscales and the total score (Table 5). The pASES-e has also demonstrated acceptable levels of relative reliability.[5,9] The ICCs for the various subscales of the English version ranged from 0.64 to 0.90,[5] whereas for the German version, higher ICCs were reported (0.74-0.93) (Table 5).[9]

Minimal detectable change and SEM

Absolute reliability has been reported only for the German version of the pASES-e (Table 5). The SEM for the satisfaction subscale was 1.62 (95% CI, −2.05 to 3.05); the function-affected subscale was 6.23 (95% CI, −15.8 to 9.88) and function-unaffected subscale was 11.62 (95% CI, −23.74 to 23.28).[9] Minimal detectable change (MDC) 95% for the pASES-e was 16.05 points.[9]

Responsiveness

Both the PREE and ASES-e have been reported to be highly sensitive to change. Usually for a scale to be sensitive to change, a score of 0.8 or above for both standardized response mean (SRM) and effect size (ES) is expected. The PREE demonstrated slightly higher ES and SRM when compared with the pASES-e (Table 6). The ES and SRM for the PREE, as reported in 3 studies, ranged from 1.7 to 1.1 and 1.3 to 1.6, respectively.[8,26,27] The ES and SRM for the pASES-e total score were 1.32 and 1.17, respectively.[27]

Only 1 study has established the longitudinal construct validity of the PREE and pASES-e.[26] The change scores of the 2 measures and DASH was used in the analysis (Table 6). For the PREE, the correlations ranged from 0.77 with the pASES-e pain scale to 0.33 with the pASES-e function scale and demonstrated a moderate correlation ($r = 0.62$) with the DASH. For the pASES-e, the pain subscale exhibited high to moderate correlations with the PREE and DASH (0.41-0.85). The function subscale correlated the least with correlations ranging from 0.23 to 0.40.

Angst et al[27] have studied the sensitivity and specificity of the 2 measures. The PREE and pASES-e demonstrated almost the same sensitivity indices (Table 6). The PREE discriminated between much better and the other categories with an area under the curve of 0.68

Table 4
Summary of validity properties, cross-cultural adaptations, and administrative and responder burden of the PREE and pASES-e

Clinical measurement property	Data extracted	
	PREE	pASES-e
Content validity (includes analysis of questions, floor and ceiling effects, missing items)	• All questions were easily understood and completed by all patients[28] • No patients had difficulty completing the questionnaire (Japanese version)[8] *Floor and ceiling effects* Pain subscale: Floor = 0%; ceiling = 15%[28] Floor = 1%; ceiling = 3% (Japanese version)[8] Floor = 0%; ceiling = 14.3% (German version)[10] Function subscale: Floor = 0%; ceiling = 4%[28] Floor = 1%; ceiling = 4% (Japanese version)[8] Floor = 0%; ceiling = 5.3% (German version)[10] Total score: Floor = 0%; ceiling = 4%[28] Floor = 1%; ceiling = 0% (Japanese version)[8] Floor = 0%; ceiling = 5.3% (German version)[10]	• All questions were easily understood and completed by all patients[28] *Floor and ceiling effects* Pain subscale: Floor = 0%; ceiling = 14%[28] Function-affected subscale: Floor = 0%; ceiling = 4%[28] Satisfaction subscale: Floor = 3%; ceiling = 43%[28] Total score: Floor = 0%; ceiling = 3%[28]
Factorial validity	• Principal component analysis with varimax rotation supported a 4-factor model with only 1 crossloading[26] • PREE pain has a unidimensional structure; PREE function has a bidimensional structure (Japanese version)[8]	Principal component analysis with varimax rotation supported a 4-factor model with 2 crossloadings[26]
Construct convergent validity	Reported high correlations (>0.70): Pain subscale: PREE function score[7,8] PREE total score[7,8,26] pASES-e pain subscale[5,26] QuickDASH[25] DASH total (Japanese version)[8] Function subscale: PREE pain score[7,8] PREE total score[7,26] DASH total score[8,10,26] QuickDASH[7,25] DASH[5] Total score: PREE pain[7,8,26] PREE function[7,8,26] pASES-e pain[26] DASH total[5,8,10,26] QuickDASH[7,25] pASES-e total score[28] Reported moderate correlations (0.30-0.70): Pain subscale: PREE function[26] pASES-e function affected[26] DASH total[10,26] SF-36 PCS score[5,10] QuickDASH[7] Function subscale: PREE pain[26] pASES-e pain[26] pASES-e function affected[26] SF-36 PCS score[5,8,10,26,28] cmASES[10] Total score: pASES-e function affected[26] SF-36 PCS score[5,8,10,26] cmASES[28]	Reported high correlations (>0.70): Pain subscale: PREE pain subscale[5,26] PREE total score[9,26] DASH[5] Function-affected subscale: DASH[9] SF-36 PCS[9] PREE function score[9] PREE total score[9] Satisfaction subscale: None Total score: DASH[9] PREE function score[9] PREE total score[9] Reported moderate correlations (0.30-0.70): Pain subscale: PREE function subscale[26] pASES-e function affected[26] DASH total score[9,26] SF-36 PCS score[5,9] cmASES[10] Function-affected subscale: PREE pain[26] PREE function[26] PREE total score[26] pASES-e pain[26] DASH total score[26] SF-36 PCS score[5,26] cmASES[10] Satisfaction subscale: SF-36 PCS score[5,10] DASH total score[10] Total score: cmASES[10,28] SF-36 PCS score[5,10,28]
Construct divergent validity	Pain subscale: SF-36 mental component summary[8] Function subscale with: SF-36 mental component summary[8,26,28] Total score with: SF-36 mental component summary[5,7,8,10,26]	Pain subscale: SF-36 mental component summary[10] Function affected subscale with: SF-36 mental component summary[10,26] Satisfaction subscale with: SF-36 mental component summary[10] Total score with: SF-36 mental component summary[10,28]
Cross-cultural adaptation	• German version 10 translated with inputs from the developer and found psychometric properties that was equivalent to the original English version • Japanese version 8 found the adapted version to exhibit clinical measurement properties equivalent to the original English version	• German version[9] found clinical measurement properties that were equivalent to the original English version

(*continued on next page*)

Table 4 (*continued*)

Clinical measurement property	Data extracted	
	PREE	pASES-e
Response burden	• Time to complete: 3 min[28] • Clarity of items: Was easily understood[27] • Reading level: Flesch level of 6.3 that is equivalent to grade 6 reading level[29]	• Time to complete: 3 min[28] • Clarity of items: Was easily understood[27]

PREE = Patient-Rated Elbow Evaluation; pASES-e = self-report section of the American Shoulder and Elbow Surgeons—elbow form; DASH = Disabilities of the Arm, Shoulder and Hand; SF-36 = 36-Item Short Form Health Survey; PCS = physical component summary; cmASES = clinical modified American Shoulder and Elbow Surgeons.

(95% CI, 0.57-0.81), and it was 0.67 (95% CI, 0.53-0.80) for the pASES-e, which supports the power of discrimination of the measure. The PREE was more specific (0.71 vs 0.69) and the pASES-e (0.65 vs 0.63) was more sensitive.[27]

Discussion

This study from its review of 9 articles concluded that strong clinical measurement properties exist for both the PREE and the

Table 5
Summary of reliability indices for the PREE and pASES-e

Clinical measurement property	Data extracted	
	PREE	pASES-e
Test-retest reliability (ICC)	Pain subscale: 0.92 (Japanese version)[8] 0.90 in RA cases (Japanese version)[8] 0.88 (95% CI, 0.78-0.94)[5] 0.73 (German version)[10] Function subscale: 0.93 (Japanese version)[8] 0.86 in RA cases (Japanese version)[8] 0.89 (95% CI, 0.79-0.94)[5] 0.82 (German version)[10] Total score: 0.94 (Japanese version)[8] 0.90 in RA cases (Japanese version)[8] 0.95 (95% CI, 0.86-0.98)[5] 0.80 (German version)[10]	Pain subscale: 0.90 (95% CI, 0.82-0.94) (German version)[9] 0.89 (95% CI, 0.80-0.94)[5] Function subscale (affected): 0.87 (95% CI, 0.78-0.92) (German version)[9] 0.79 (95% CI, 0.66-0.88)[5] Function subscale (unaffected): 0.74 (95% CI, 0.53-0.87) (German version)[9] 0.64 (95% CI, 0.42-0.78)[5] Satisfaction subscale: 0.84 (95% CI, 0.71-0.91)[5] Total score (affected): 0.93 (95% CI, 0.88-0.96) (German version)[9] Total score (unaffected): 0.92 (95% CI, 0.84-0.96) (German version)[9]
SEM	None	Pain subscale: 8.94 (95% CI, −21.23 to 14.74) (German version)[9] Function subscale (affected): 9.04 (95% CI, −20.83 to 15.40) (German version)[9] Function subscale (unaffected): 11.62 (95% CI, −23.74 to 23.28) (German version)[9] Satisfaction subscale: 1.32 (95% CI, −2.05 to 3.25) (German version)[9] Total score (affected): 6.23 (95% CI, −15.8 to 9.88) (German version)[9] Total score (unaffected): 6.61 (95% CI, −14.70 to 12.02) (German version)[9]
$MDC_{95\%}$	None	Pain subscale: 23.51 (German version)[9] Function subscale (affected): 23.42 (German version)[9] Function subscale (unaffected): 31.43 (German version)[9] Satisfaction subscale: 3.85 (German version)[9] Total score (affected): 16.05 (German version)[9] Total score (unaffected): 18.27 (German version)[9]
Mean retest difference (Bland and Altman plot)	None	Total score: affected side mean difference and narrowest 95% CI −2.6 ± 16.05 (German version)[9]
Internal consistency (Cronbach alpha)	Pain subscale: 0.92 (95% CI, 0.88-0.93) (Japanese version)[8] 0.93 (German version)[10] Function subscale: 0.97 (95% CI, 0.97-0.97) (Japanese version)[8] 0.95 (German version)[10] Total score: 0.95 (95% CI, 0.94-0.96)[26] 0.97 (95% CI, 0.97-0.97) (Japanese version)[8] 0.96 (German version)[10]	Pain subscale: 0.91 (German version)[9] Function subscale (affected): 0.93 (German version)[9] Function subscale (unaffected): 0.94 (German version)[9] Total score (affected): 0.93 (95% CI, 0.91-0.94)[26] 0.90 (German version)[9] Total score (unaffected): 0.90 (German version)[9]

PREE = Patient-Rated Elbow Evaluation; pASES-e = self-report section of the American Shoulder and Elbow Surgeons—elbow form; ICC = intraclass correlation coefficient; RA = rheumatoid arthritis; 95% CI, 95% confidence interval; SEM = standard error of measurement; $MDC_{95\%}$ = minimal detectable change.

Table 6
Summary of responsiveness properties of the PREE and pASES-e

Clinical measurement property	Data extracted	
	PREE	pASES-e
ES	Pain subscale: 1.7[27] 1.4 (95% CI, 1.2-1.6)[26] 1.32 (Japanese version)[8] Function subscale: 0.99[27] 1.4 (95% CI, 1.3-1.6)[26] 0.86 (Japanese version)[8] PREE total: 1.5[27] 1.7 (95% CI, 1.5-1.8)[26] 1.12 (Japanese version)[8]	Pain subscale: 1.55[27] 1.4 (95% CI, 1.2-1.6)[26] Function subscale: 0.77[27] 1.3 (95% CI, 1.1-1.5)[26] Total score: 1.32[27]
SRM	Pain subscale: 1.27[27] 1.3 (95% CI, 1.0-1.4)[26] 1.31 (Japanese version)[8] Function subscale: 0.97[27] 1.5 (95% CI, 1.4-1.7)[26] 1.02 (Japanese version)[8] PREE total: 1.37[27] 1.6 (95% CI, 1.4-1.8)[26] 1.28 (Japanese version)[8]	Pain subscale: 1.15[27] 1.2 (95% CI, 1.0-1.4)[26] Function subscale: 0.75[27] 1.1 (95% CI, 0.9-1.3)[26] Total score: 1.17[27]
Longitudinal validity correlation with other change scores	Pain subscale change score: With pASES-e pain: 0.85 ($P < .01$)[26] With pASES-e function: 0.24 ($P < .01$)[26] With DASH: 0.41 ($P < .01$)[26] Function subscale change score: With pASES-e pain: 0.41 ($P < .01$)[26] With pASES-e function: 0.40 ($P < .01$)[26] With DASH: 0.62 ($P < .01$)[26] PREE total change score: With pASES-e pain: 0.74 ($P < .01$)[26] With pASES-e function: 0.33 ($P < .01$)[26] With DASH: 0.62 ($P < .01$)[26]	Pain subscale change score: With PREE pain: 0.85 ($P < .01$)[26] With PREE function: 0.41 ($P < .01$)[26] With PREE total score: 0.74 ($P < .01$)[26] With DASH: 0.46 ($P < .01$)[26] Function subscale change score: With PREE pain: 0.24 ($P < .01$)[26] With PREE function: 0.40 ($P < .01$)[26] With PREE total score: 0.33 ($P < .01$)[26] With DASH: 0.23 ($P < .01$)[26]
AUC	Pain subscale: 0.65 (95% CI, 0.51-0.78); sensitivity = 0.67; specificity = 0.71[27] Function subscale: 0.62 (95% CI, 0.48-0.75); sensitivity = 0.83[a]; specificity = 0.47[27,a] Total score: 0.68 (95% CI, 0.57-0.81); sensitivity = 0.63; specificity = 0.71[27]	Pain subscale: 0.64 (95% CI, 0.51-0.78); sensitivity = 0.64; specificity = 0.71[27] Function subscale: 0.62 (95% CI, 0.47-0.78); sensitivity = 0.58[a]; specificity = 0.63[27,a] Total score: 0.67 (95% CI, 0.53-0.80); sensitivity = 0.65; specificity = 0.69[27]

PREE = Patient-Rated Elbow Evaluation; pASES-e = self-report section of the American Shoulder and Elbow Surgeons—elbow form; ES= effect size; 94% CI = 95% confidence interval; SRM = standardized response mean; DASH = Disabilities of the Arm, Shoulder and Hand questionnaire; AUC = area under the curve.
[a] Not significant.

pASES-e and that one cannot be recommended over the other. Despite the strong measurement properties reported, we identified gaps in the current evidence for both the ASES-e and the PREE. The studies reviewed were of high quality, which indicates that the recommendations that are generated from this review are based on high-quality clinical measurement research.

Content validity was reported as being acceptable for both PROMs as multiple studies suggested that all items were easy to understand. However, the content of these PROMs has not been adequately analyzed using formal content analyses strategies. This is not unexpected because many PROMs have not used rigorous and structured content analysis during development or evaluation. However, the content of these questionnaires has been analyzed against the international standard setup by the World Health Organization called the International Classification of Functioning, Disability and Health.[31-33] Other methods that can provide a structured approach to evaluating content such as cognitive interviews, qualitative studies, concept mapping, and others have also not been reported as ways of analyzing the content of either measure. Despite this lack of formal evaluation of content, face validity suggests that they both measure pain and disability in patients with elbow conditions.

Floor and ceiling effects in a questionnaire can pose potential problems because when patients score at the extremes of a scale, further worsening or improvement may not be detected. A floor or ceiling effect has been defined as being problematic if more than 15% of the participants are at either ends of the spectrum[34]; however, this was not a problem for the PREE and pASES-e, except for the satisfaction section for the pASES-e. It has been previously argued that satisfaction is a concept that is more irrelevant to the constructs of pain and disability[35,36] and should be considered as a separate section, and efforts should not be made to include the score from the satisfaction section when total score calculation is attempted for pASES-e.[26]

Unidimensionality is another important clinical measurement property of outcome measures because measures are typically

defined as measuring a specific construct and should be able to measure that as a single trait.[37] Exploratory factor analyses of both the PREE and the pASES-e suggested that they were not unidimensional. However, internal consistency was high (>0.90) for both the PREE and the pASES-e supporting unidimensionality. These findings need to be explored further using powerful statistical methods like confirmatory factor analysis, structural equation modeling, or Rasch analysis.[38-41] These confirmatory analyses will determine the dimensionality of these measures and may suggest amendments that may render these measures unidimensional if they are not.[37,40]

The PREE and pASES-e have demonstrated expected correlations (>0.70) with other measures measuring a similar construct like the DASH, QuickDASH, and between themselves. This indicates that these scales are measuring similar constructs and can be used solely based on the clinical need and ease of use. Both these measures have demonstrated moderate correlations with the physical component summary score of the SF-36, underlining the effect of disability on overall physical health status. There is little evidence in the literature surrounding the known group validity of these 2 measures making it difficult to make interpretations in specific populations. Future studies are needed to assess the validity of these measures on using additional clinical populations.

Both these measures have been deemed to be fit for use in either individual assessments or in a group as they have demonstrated high test-retest reliability (ICC >0.90). However, these indices are measures of relative reliability and not absolute reliability. For a measure to be clinically useful, measures of absolute reliability such as SEM and MDC should be provided.[21] These indices create estimates that are free of measurement error and capture changes that are important clinically and for the patients themselves. SEM and MDC have been calculated only for the German version of the pASES-e.[9] Another measure of reliability that was reported for both the PROMS is the Bland and Altman technique; it is used to assess reliability and can detect bias between assessments.[42] Future studies should focus on deriving these measures to enhance clinical applicability by being able to detect and quantify change as it occurs.

The PREE and pASES-e are sensitive to change established by large ES and SRM from multiple studies.[8,26,27] Therefore, these 2 measures can detect true change in the patient's status as it happens.[43] SRM and ES are distribution-based methods and can only capture real change.[44,45] To capture the clinically important change, future studies need to be conducted that use anchor-based methods, such as the global rating of change to calculate minimal clinically important difference (MCID).[21,43,45] This is a vital characteristic of any measure that would help increase the clinical utility of tools by increasing the confidence of the clinician or researcher when interpreting the results of an outcome measure.

A report on the quality assessment tools used to evaluate the quality of systematic reviews identified 86 different tools; they recommend that these tools should be developed using rigorous processes, based on evidence, and should be reliable.[46] A fundamental component of systematic reviews is that they assess quality of the study or risk of bias. The 2 most commonly used appraisal tools in clinical measurement research are the QA-CMS[16] and the COnsensus-based Standards for the selection of health Measurement INstruments criteria. Previous studies have found that the QA-CMS has demonstrated excellent test-retest reliability (ICC >0.90), and this is consistent with the findings of this study.[17,18,47] The COnsensus-based Standards for the selection of health Measurement INstruments was developed with an international committee and has gained acceptance for use in measurement reviews, although reliability has not been consistently good.[48] The QA-CMS was selected for its ease of use and track record of reliability and flexibility to deal with different types of measurement studies.

The strengths of the present study are the inclusion of articles from any language; to reduce the potential for bias because measures may perform differently across contexts. Given that one of our main findings was that there is insufficient volume of clinical measurement evidence to fully define all the measurement properties of these tools, it was critical that we did not miss relevant studies. Our review had limitations that limit the recommendations arising from this work, and it should be considered when interpreting our findings. We could not calculate a pooled estimate for key psychometric properties because of the small number of studies and the heterogeneity in the studies. One of the coauthors of the article is the developer of the PREE and our critical appraisal tool that may have affected our reliability. However, this author was secondarily involved, and primary data extraction was performed by the first and last authors. Because our findings were equally supportive of the PREE and pASES-e, there does not appear to be a bias in our findings or recommendations.

Future studies investigating the clinical measurement properties of the PREE and pASES-e may consider the following recommendations in their design and reporting: use powerful statistical methods that are part of modern psychometric methods such as Rasch analysis to analyze the fit of the PROMs to the model requirements; provide clear justification of the sample size used; state clear hypotheses of what is expected for the properties being tested; calculate clinically important estimates like MCID, SEM, and others; and provide clear and specific conclusions that reflect the purpose of the review.

In conclusion, the PREE and pASES-e have been shown to be valid, reliable, and sensitive to change in both clinical and research settings based on high-quality evidence. However, to improve the clinical utility, clinically important measurement estimates such as the SEM and MCID should be established. Future studies should focus on using modern psychometric methods to evaluate these 2 PROMs against rigorous statistical models.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jht.2018.07.004.

References

1. Danis M, Solomon M. Providers, payers, the community, and patients are all obliged to get patient activation and engagement ethically right. *Health Aff*. 2013;32(2):401–407.
2. Small N, Bower P, Chew-Graham CA, Whalley D, Protheroe J. Patient empowerment in long-term conditions: development and preliminary testing of a new measure. *BMC Health Serv Res*. 2013;13:263.
3. Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *J Clin Epidemiol*. 1992;45(7):743–760.
4. MacDermid JC. Patient-rated elbow evaluation. Available at: http://www.srs-mcmaster.ca/Portals/20/pdf/research_resources/PREE.pdf. Accessed October 9, 2011.
5. MacDermid JC. Outcome evaluation in patients with elbow pathology: issues in instrument development and evaluation. *J Hand Ther*. 2001;14(2):105–114.
6. King GJ, Richards RR, Zuckerman JD, et al. A standardized method for assessment of elbow function. Research Committee, American Shoulder and Elbow Surgeons. *J Shoulder Elbow Surg*. 1999;8(4):351–354.
7. Liem IS, Kolling C, Marks M, Nelissen RG, Goldhahn J. Development of a score set to measure function and quality of life in patients suffering from elbow pathology. *Arch Orthop Trauma Surg*. 2012;132(6):831–837.
8. Hanyu T, Watanabe M, Masatomi T, et al. Reliability, validity, and responsiveness of the Japanese version of the patient-rated elbow evaluation. *J Orthop Sci*. 2013;18(5):712–719.
9. John M, Angst F, Awiszus F, King GJ, MacDermid JC, Simmen BR. The American Shoulder and Elbow Surgeons Elbow Questionnaire: cross-cultural adaptation into German and evaluation of its psychometric properties. *J Hand Ther*. 2010;23(3):301–313. quiz 314.

10. John M, Angst F, Pap G, Junge A, Mannion AF. Cross-cultural adaptation, reliability and validity of the Patient Rated Elbow Evaluation (PREE) for German-speaking patients. *Clin Exp Rheumatol*. 2007;25(2):195–205.
11. The B, Reininga IH, El Moumni M, Eygendaal D. Elbow-specific clinical rating systems: extent of established validity, reliability, and responsiveness. *J Shoulder Elbow Surg*. 2013;22(10):1380–1394.
12. Nourbakhsh E, Nugent R, Wang H, Cevik C, Nugent K. Medical literature searches: a comparison of PubMed and Google Scholar. *Health Info Libr J*. 2012;29(3):214–222.
13. Shariff SZ, Bejaimal SA, Sontrop JM, et al. Retrieving clinical evidence: a comparison of PubMed and Google Scholar for quick clinical searches. *J Med Internet Res*. 2013;15(8):e164.
14. MacDermid JC. The Patient Rated Elbow Evaluation (PREE)—user manual. Available at: http://www.srs-mcmaster.ca/Portals/20/pdf/research_resources/PREE%20User%20Manual_June%202010.pdf. Accessed June 20, 2014.
15. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700.
16. MacDermid J. Quality appraisal for clinical measurement studies—evaluation form and guidelines. In: Law MC, MacDermid J, eds. *Evidence-Based Rehabilitation: A Guide to Practice*. 3rd ed. Thorofare, NJ: Slack; 2014:325–338.
17. Roy JS, Desmeules F, MacDermid JC. Psychometric properties of presenteeism scales for musculoskeletal disorders: a systematic review. *J Rehabil Med*. 2011;43(1):23–31.
18. Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum*. 2009;61(5):623–632.
19. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257–268.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
21. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol*. 1992;45(12):1341–1345.
22. Grayson K, Rust R. Interrater reliability. *J Consumer Psychol*. 2001;10(1-2):71–73.
23. Wong SP, McGraw KO. Confidence intervals and F tests for intraclass correlations based on three-way random effects models. *Educ Psychol Meas*. 1999;59(2):270–288.
24. Wright BD. Interpreting reliabilities. *Rasch Meas Trans*. 1998;11(4):602.
25. Schmidt T, Stangl R. Indications, complications and results of elbow arthroscopy: a retrospective assessment with correlation analysis of the German version of the patient rated elbow evaluation (PREE) score. *Obere Extrem*. 2013;8(4):220.
26. Vincent JI, King GJW, Grewal R. Validity and sensitivity to change of patient-reported pain and disability measures for elbow pathologies. *J Orthop Sports Phys Ther*. 2013;43(4):263–274.
27. Angst F, Goldhahn J, Drerup S, et al. Responsiveness of five outcome measurement instruments in total elbow arthroplasty. *Arthritis Care Res*. 2012;64(11):1749–1755.
28. Angst F, John M, Pap G, et al. Comprehensive assessment of clinical outcome and quality of life after total elbow arthroplasty. *Arthritis Care Res*. 2005;53(1):73–82.
29. Kincaid JP, Fishburne Jr RP, Rogers RL, et al. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy-enlisted personnel. *Res Branch Rep*. 1975;8(75):1–39.
30. Vincent JI, MacDermid JC, King GJ. Response to: critical assessment of patient-reported outcome measures. *J Orthop Sports Phys Ther*. 2013;43(7):513–514.
31. World Health Organization. Towards a common language for Functioning, Disability and Health, ICF. Available at: http://www.who.int/classifications/icf/training/icfbeginnersguide.pdf. Accessed June 20, 2014.
32. Vincent JI, MacDermid JC, King GJ, et al. Linking of the Patient Rated Elbow Evaluation (PREE) and the American Shoulder and Elbow Surgeons–elbow questionnaire (pASES-e) to the International Classification of Functioning Disability and Health (ICF) and hand core sets. *J Hand Ther*. 2015;28(1):61–68.
33. Naughton N, Algar L. Linking commonly used hand therapy outcome measures to individual areas of the International Classification of Functioning: a systematic review. *J Hand Ther*. 2018. https://doi.org/10.1016/j.jht.2017.11.039.
34. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4(4):293–307.
35. Chung KC, Hamill JB, Kim HM, Walters MR, Wilkins EG. Predictors of patient satisfaction in an outpatient plastic surgery clinic. *Ann Plast Surg*. 1999;42(1):56–60.
36. Roy JS, Macdermid JC, Goel D, Faber KJ, Athwal GS, Drosdowech DS. What is a successful outcome following reverse total shoulder arthroplasty? *Open Orthop J*. 2010;4:157–163.
37. Linacre JM. Detecting multidimensionality: which residual data-type works best? *J Outcome Meas*. 1998;2(3):266–283.
38. Smith Jr EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas*. 2002;3(2):205–231.
39. Andrich D. *Rasch Models for Measurement*. Vol. 07-068. Newbury Park, CA: Sage Publications; 1988:95.
40. DeMars C. *Item Response Theory*. Oxford, New York: Oxford University Press; 2010:131.
41. Brown TA. *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Publications, Incorporated; 2006.
42. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2):135–160.
43. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials*. 1991;12(4 suppl l):142S–158S.
44. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care*. 1989;27(3 suppl l):S178–S189.
45. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis*. 1987;40(2):171–178.
46. Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol*. 2007;36(3):666–676.
47. MacDermid JC, Walton DM, Avery S, et al. Measurement properties of the neck disability index: a systematic review. *J Orthop Sports Phys Ther*. 2009;39(5):400–417.
48. Mokkink LB, Terwee CB, Gibbons E, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. *BMC Med Res Methodol*. 2010;10:82.

JHT Read for Credit
Quiz: # 610

Record your answers on the Return Answer Form found on the tear-out coupon at the back of this issue or to complete online and use a credit card, go to *JHTReadforCredit.com*. There is only one best answer for each question.

#1. The study design is
   a. qualitative
   b. a case series
   c. a systematic review
   d. RCTs

#2. The two elbow PROMs which were considered were the _____ and the _____
   a. pAES-e, PREE
   b. DASH, short DASH
   c. EES, SEE
   d. MMT, UES

#3. Both PROMs were shown to be
   a. valid
   b. sensitive to change
   c. reliable
   d. all of the above

#4. A report on the quality assessment tools used to evaluate the quality of systematic reviews identified approximately _____ different tools
   a. 25
   b. 75
   c. 85
   d. 105

#5. The SEM of both PROMs has been firmly established
   a. true
   b. false

When submitting to the HTCC for re-certification, please batch your JHT RFC certificates in groups of 3 or more to get full credit.