JSES Open Access 3 (2019) 213–218

Contents lists available at ScienceDirect

JSES Open Access

journal homepage: www.elsevier.com/locate/jses

The Dutch version of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form is a reliable and valid questionnaire for shoulder problems

Quinten T.M. Felsch, MD [a,b,*], Peter Sievert, MD [a], Martijn G.M. Schotanus, MSc, PhD [a], Edwin J.P. Jansen, MD, PhD [a]

[a] Department of Orthopedic Surgery, Zuyderland Medical Center, Sittard-Geleen, the Netherlands
[b] Department of Sports Medicine, Schulthess Klinik, Zörich, Switzerland

ARTICLE INFO

Keywords:
ASES
cross-cultural adaptation
validation
Dutch
PROM
questionnaire

Level of evidence: Basic Science, Validation of Outcomes Instrument

Background: The self-assessment section of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASESq) is one of the most used patient-reported outcome measures for general shoulder problems. This study was performed to establish a valid Dutch version of the ASESq (ASESq-NL).

Materials and Methods: A clinical prospective, nonrandomized study was performed. Translation of the ASESq into Dutch was done following the guidelines of cross-cultural adaptation. Patients older than 17 years of age with shoulder problems were included. Patients who declined to participate or insufficiently completed questionnaires were excluded. For test-retest reliability analysis, the intraclass correlation coefficient (ICC) was calculated and an interval of 7-28 days between test and retest was set. Cronbach alpha was used to determine internal consistency. Dutch validated versions of the Shoulder Pain and Disability Index (SPADI) and 36-Item Short Form Health Survey (SF-36) were completed and compared with the ASESq-NL to evaluate construct validity using a Spearman rank correlation coefficient calculation.

Results: A total of 92 patients were included. Test-retest reliability was excellent with an ICC of 0.82. The mean test-retest interval was 13 days (standard deviation 4.4). Internal consistency was good, with a Cronbach alpha of 0.83. Construct validity of the ASES questionnaire was good. All domains of the ASESq-NL had significant ($P < .05$) correlations with the domains of the SPADI and the SF-36, except for the SF-36 domains stability with “physical function and energy" and "emotional well-being."

Conclusion: The Dutch ASES questionnaire is a valid and reliable tool for the evaluation of shoulder problems and is permissible for implementation into the Dutch health care system.

With health care costs on the rise, national governments and health insurances increasingly demand the evaluation of treatment regimens with valid outcome measures to assess implementation and financial approval to the health care system.[14,17]

The use of validated outcome instruments is widely accepted for the assessment of patient outcome to advance medical treatment in the clinical and research setting.[18,24]

Outcome measures can be generally divided into clinician-reported and patient-reported outcome instruments. Clinician-reported measures focus on objective outcome assessed by a trained health care professional. In contrast, patient-reported outcome measures (PROMs) specifically evaluate the patient's perspective on health improvement; as a result, PROMs have gained widespread recognition.[23] A range of PROMs exist, from general health questionnaires, like the 36-Item Short Form Health Survey (SF-36), to more disease- and joint-specific questionnaires, like the patient self-evaluation section of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASESq).[24]

The ASESq is one of the most used PROMs for shoulder problems.[7,24,26] It is easy and quick to use and validated for the evaluation of shoulder problems.[20] The ASESq has a broad scientific background, with citations in more than 1000 papers and a good scoring of the AO Handbook on Musculoskeletal Outcome Measures and Instruments.[27,30] Schmidt et al conducted a standardized and systematic review of 11 shoulder-specific PROMs using the

Institutional Review Board approval was received approval was received from the medical ethical committee METC-ZUYD (study no. 13-N-122).

* Quinten T.M. Felsch, MD, Department of Sports Medicine, Schulthess Klinik, Lengghalde 6, 8008 Zürich, Switzerland.
E-mail address: quinten.felsch@kws.ch (Q.T.M. Felsch).

https://doi.org/10.1016/j.jses.2019.06.002

Evaluating the Measurement of Patient-Reported Outcomes evaluation tool for PROMs.[28,32] The ASESq obtained the best overall score and was consecutively among the top 3 outcome measures in this study.[28] The ASESq has been translated and validated into many languages, including German, Italian, Spanish, Finnish, Portuguese, Turkish and Arabic.[8,12,15,21,22,34,35] A validated Dutch version is currently not available. The multilingual availability of the ASESq makes it particularly suited for clinical research as well for the use in multicultural societies like the Netherlands. This clinical prospective, nonrandomized study was performed in order to cross-culturally adapt and validate the Dutch version of the ASESq (ASESq-NL).

Materials and methods

Questionnaires

The American Shoulder and Elbow Surgeons Standardized Assessment Form was developed during 1990 to 1993 by the ASES to address the need for a state-of-the-art assessment tool for all shoulder patients regardless of diagnosis. The American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form consists of 3 sections: demographic information, patient self-evaluation (ASESq), and physician assessment.[24] We focused on the ASESq, as a PROM, together with the ASES shoulder score index, which is calculated from the items of the ASESq.[8,12,15,21,22,34,35] The ASESq contains 18 questions divided over 3 sections: pain, instability and activities of daily living (ADL). The ASES shoulder score index is derived from the visual analog scale (VAS) for pain and the cumulative ADL score. The ADL score consists of 10 questions that assess ADL for both shoulders, graded on a 4-point ordinal scale, from 0 (unable to do) to 3 (not difficult). The shoulder score index (X) has a range from 0 (most disability) to 100 (least disability) and can be calculated with the formula: $X = [(10 - \text{VAS pain score}) \times 5] + [(5/3) \times \text{cumulative ADL score}]$.

The validation of the ASESq-NL is achieved by comparing its 3 domains “pain”, “stability” and “activity of daily living” with the corresponding 8 domains of the 36-Item SF-36 and the 2 domains of the Shoulder Pain and Disability Index (SPADI).[10,11,25] The SF-36 is one of the most used generic patient-reported health measures, but is also specifically validated for shoulder complaints.[6] The SPADI is a tested and widely used self-assessment instrument for the shoulder.[7,25] Both measures were adapted and validated for Dutch language and have both been used for the outcome validation of the original ASESq as well as for the crosscultural adaptation and validation of the ASESq into other languages.[8,12,22,24,30,34,35]

Translation and linguistic validation method

A forward-backward translation protocol according to the guidelines of cross-cultural adaptation was used following the 5 steps of translation, synthesis, back-translation, expert committee review and pretesting.[1,13] Forward translation was done separately by 2 native Dutch speakers, one acting as an informed translator (orthopedic resident) and the other as an uninformed translator (medical student). A Dutch consensus version of the ASESq was created and checked for cross-cultural differences. Three cross-cultural dissimilarities were identified. Question 4 of the ASESq refers to pain medication and provides example substances aspirin, Advil and Tylenol. The example medication is not typical for the Netherlands and was replaced with paracetamol and ibuprofen, both comparable substances commonly used in the Netherlands. In question 5, narcotic pain medication was translated to medication requiring a doctor's prescription, with the narcotic medication tramadol and codeine as examples. Both questions are not part of the shoulder score index. In the self-evaluation section concerning activities of daily living, the weight in question 7, “lift 10 lbs above the shoulder,” was metrically converted to 5 kg. After completing the Dutch consensus version of the ASESq, a backward translation by a native English speaker not working in the medical field was executed. Both versions were reviewed by an expert committee. Forward and backward translations revealed no severe differences or language difficulties. After approval of the preliminary ASESq-NL a pretest was performed on 20 patients to reveal any problems in handling and understanding. No significant difficulties were reported. As a result, the final Dutch ASES questionnaire (ASESq-NL) was concluded.

Study population

The study was conducted by the Department of Orthopedic Surgery in the Zuyderland Medical Center in the Netherlands. Between October and December 2013, all patients older than 17 years of age who were referred to our clinic with shoulder problems were asked to participate. Patients who delivered incomplete questionnaires or denied participation were excluded. A confirmation letter of the appointment at our outpatient clinic was delivered to the participant by post accompanied by the first set of questionnaires (ASESq-NL, SPADI, SF-36). Participants were instructed to complete the questionnaires unassisted. For test-retest analysis, a second set of questionnaires (ASESq-NL, SPADI, SF-36) was completed at our outpatient clinic before visiting the orthopedic surgeon. Questionnaires with a test-retest interval of less than 1 or more than 4 weeks were excluded, as suggested in the literature.[1,3,29,33] Patients who received an intervention during the test-retest interval were also excluded. Interventions were defined as shoulder injections or operations, not counting oral pain medication.

Assessment of psychometric properties

Reliability and validity were assessed to determine the quality of the measurement instrument.

Reliability

Reliability refers to the degree to which results of an instrument can be replicated on recurring measurements across time (test-retest) and among related items on the instrument (internal consistency). It can be expressed by a value from 0, no reliability, to 1, absolute reliability.[29]

The Cronbach alpha was used to calculate internal consistency. It is a widely accepted tool for homogeneity calculation and has been used in most comparable ASESq validation studies.[12,33] Values greater than 0.70 reflected a sufficient correlation between the items of a questionnaire. A result between 0.70 and 0.79 was considered as fair, between 0.80 and 0.89 as good and $\geq$0.90 as excellent internal consistency.[9] A Cronbach alpha above 0.90 could however imply that items on an instrument are too homogenous and thus redundant. Generally, a maximum alpha value of 0.90 has been recommended.[31]

Test-retest reliability is based on the assumption that 2 separate measurements should be the same if no change occurred. A fitting time interval between test and retest is crucial. A short interval could lead to a memory bias and a long interval to an actual change of status. Following the guidelines of cross-cultural adaptation, a test-retest analysis on 30 or more patients with a time interval from 1 to 4 weeks was performed. The intraclass

Table I
Patient characteristics

	No. (%)
Sex	
Female	47 (51)
Age, yr, mean (SD)	55 (±12)
Affected side	
Right shoulder	57 (62)
Left shoulder	29 (31.5)
Bilateral	6 (6.5)
Diagnosis	
Tendinitis calcarea	14 (15)
Subacromial impingement	33 (36)
Biceps tendinitis	5 (5)
AC osteoarthritis	9 (10)
SC dislocation	1 (1)
Rotator cuff syndrome	2 (2)
Rotator cuff rupture	8 (9)
Labrum lesion	3 (3)
Multidirectional instability	1 (1)
Glenoid fracture	1 (1)
Omarthrosis	2 (2)
Frozen shoulder	11 (12)
Unknown	2 (2)

SD, standard deviation; *AC*, acromioclavicular; *SC*, sternoclavicular

correlation coefficient (ICC) was calculated to evaluate test-retest reliability. It is a widely accepted tool for the evaluation of test-retest reliability.[16] An ICC of 0 indicated no agreement, and an ICC of 1 absolute agreement, between test and retest. An ICC within 0.60-0.74 was considered good and an ICC >0.74 was considered excellent.[9] Visual presentation of the test-retest reliability was done with a Bland-Altman plot,[5] which allows more insight into the spread of data for test-retest analysis.

Validity

Validity is the degree to which an instrument measures what it is supposed to measure. The 3 domains of the ASESq-NL were compared with the corresponding domains of the SF-36 and SPADI to evaluate construct validity. This was done using the Spearman rank correlation coefficient. Statistical analysis was performed using SPSS 24.0 (IBM, Armonk, NY, USA) and Excel 2010 (Microsoft, Redmond, WA, USA), considering a $P \leq .05$ as significant.

Results

A total of 103 patients were asked to participate. Two patients refused to participate and 9 patients did not complete >70% of the questionnaires. The remaining 92 patients were included. The patient characteristics are shown in Table I.

Table II
Test-retest ASESq

ASESq domains (n = 37)	ICC (95% CI)	*P* value
Pain	0.80 (0.61, 0.90)	<.01
Stability	0.77 (0.55, 0.88)	<.01
Daily activities	0.84 (0.68, 0.92)	<.01
ASESq score total	0.82 (0.65, 0.91)	<.01

ASESq, American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form; *ICC*, intraclass correlation; *CI*, confidence interval.
All domains had significant ICCs. The domain Daily activities had the highest correlation and Stability the lowest correlation.

Internal consistency and test-retest reliability

Internal consistency was good, with a Cronbach alpha of 0.83. According to the guidelines of cross-cultural adaptation, a minimum sample size of 30 is required for test-retest reliability analysis. Test-retest was performed in 37 patients. The mean interval between test and retest was 13 days (standard deviation 4.4), with an excellent ICC of 0.82 (95% confidence interval 0.65, 0.91; $P < .01$) for the total ASESq score. The ICC for the subgroups of the ASESq-NL are shown in Table II. The Bland-Altman plot (Fig. 1) shows good agreement between test and retest, which is consistent with the calculated ICCs.

Construct validity

All domains of the ASESq-NL had significant correlations ($P < .05$) with the domains of the SPADI and the SF-36 (see Tables III and IV), except for the 2 SF-36 domains "stability with physical function and energy" and "emotional well-being." For the ASESq-NL domain "stability," only 84 patients completed the stability questions and were available for analysis.

Discussion

The most important finding of this study is that the Dutch version of the ASESq is a valid and reliable tool for assessing shoulder problems.

No severe problems were encountered during the translation and adaptation process. A good internal consistency with a Cronbach alpha of 0.83 was found. Alpha values between 0.70 and 0.80 are regarded as acceptable for the comparison of groups.[4] During comparison of our Cronbach alpha with the alpha values of previous adaptation studies, similar results were found, with the

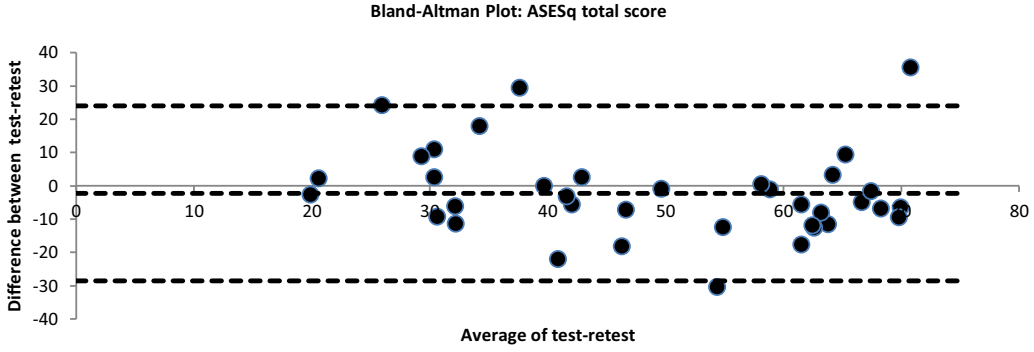


Figure 1 Bland-Altman plot. Top horizontal line: upper limit of agreement, 24; mid horizontal line: mean (bias), −2.28; lower horizontal line: lower limit of agreement, −29. *ASESq*, American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form.

Addendum ASESq-NL: Dutch ASES Shoulder Form

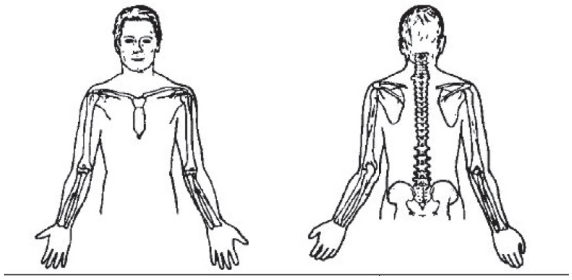
ASES-schouder vragenlijst

Nederlandse versie

De ASES Score is een zelfevaluatie door de patiënt. Beantwoord hieronder de vragen.

ASES zelf-evaluatie: Pijn

1. Heeft u pijn in uw schouder? Ja / Nee
2. Waar heeft u pijn in uw schouder?



3. Heeft u 's nachts pijn in uw schouder? Ja / Nee
4. Gebruikt u pijnstillers (paracetamol, ibuprofen, etc.)? Ja / Nee
5. Gebruikt u recept-plichtige pijnstillers (tramadol, codeine of sterker)? Ja / Nee
6. Hoeveel tabletten voor pijnstilling gebruikt u gemiddeld per dag? Aantal: .....
7. Geef op de lijn aan hoeveel pijn u vandaag heeft?

Geen ---------- Maximaal

ASES zelf-evaluatie: Instabiliteit

1. Voelt uw schouder instabiel aan (alsof hij uit de kom gaat)? Ja / Nee
2. Geef op de lijn aan hoe instabiel uw schouder is?

Geen ---------- Maximaal

ASES zelf-evaluatie: ADL

Geef aan d.m.v. omcirkelen hoe goed u volgende activiteiten kunt uitvoeren.

0 = Kan ik niet

1 = Zeer moeilijk

2 = Iets moeilijk

3 = Niet moeilijk

	Rechts	Links
1. Een jas aandoen?	**0 - 1 - 2 - 3**	**0 - 1 - 2 - 3**
2. Slapen op de aangedane zijde?	**0 - 1 - 2 - 3**	**0 - 1 - 2 - 3**
3. De rug wassen/ bh-bandje achteren sluiten?	**0 - 1 - 2 - 3**	**0 - 1 - 2 - 3**
4. Toiletgang?	**0 - 1 - 2 - 3**	**0 - 1 - 2 - 3**
5. Haar kammen?	**0 - 1 - 2 - 3**	**0 - 1 - 2 - 3**
6. Hoge plank bereiken?	**0 - 1 - 2 - 3**	**0 - 1 - 2 - 3**
7. 5 Kg boven schouderhoogte tillen?	**0 - 1 - 2 - 3**	**0 - 1 - 2 - 3**
8. Bovenhands een bal gooien?	**0 - 1 - 2 - 3**	**0 - 1 - 2 - 3**
9. Normale werk doen?	**0 - 1 - 2 - 3**	**0 - 1 - 2 - 3**
10. Normale sport doen?	**0 - 1 - 2 - 3**	**0 - 1 - 2 - 3**

Figure 2 ASESq-NL: Dutch ASES shoulder form. *ASESq*, American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form.

exception of the Cronbach alpha described by Goldhahn et al.[8,12,20–22,35] An explanation could be that Goldhahn et al only included patients who underwent a primary total shoulder arthroplasty. Different alpha values could also be due to differences in cohort size, as well as possible variations in the calculation of Cronbach alpha. Considering our good and comparably rated Cronbach alpha, we are confident that a good internal consistency of the Dutch ASES questionnaire was achieved.

Furthermore, an excellent test-retest reliability was found for the Dutch ASES questionnaire with an ICC of 0.82 and a corresponding Bland-Altman plot. Compared with the studies of Celic and Goldhahn, who found an ICC of 0.94 and 0.93, respectively, the

Table III
Correlation of ASESq domains compared with the domains to the domains of the SF-36 questionnaire

SF-36	Physical function	Role limitations—physical	Role limitations—emotional	Energy /Fatigue	Emotional well-being	Social functioning	Pain
ASESq							
Activity (n = 92)							
Correlation	0.49*	0.49*	0.36*	−0.083	0.17	0.44*	0.57*
Significance (2-tailed)	0.000	0.000	0.000	0.43	0.11	0.000	0.003
Pain (n = 92)							
Correlation	−0.32*	−0.35*	−0.25*	0.082	−0.13	−0.44*	−0.61*
Significance (2-tailed)	0.002	0.001	0.015	0.44	0.24	0.000	0.000
Stability (n = 84)							
Correlation	−0.21	−0.22†	−0.22†	0.086	−0.13	−0.30*	−0.35*
Significance (2-tailed)	0.054	0.050	0.041	0.44	0.24	0.005	0.001
Total (n = 92)							
Correlation	0.53*	0.52*	0.35*	−0.11	0.16	0.50*	0.73*
Significance (2-tailed)	0.000	0.000	0.001	0.32	0.13	0.000	0.000

ASESq, American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form; *SF-36*, 36-Item Short Form Health Survey.
The strongest correlation was between ASESq—total and SF-36 Pain, whereas no significant correlation was found for the domains Energy/Fatigue and Emotional well-being, as well as between ASESq stability with physical functioning.
* Correlation was significant at the alpha ≤0.01 level (2-tailed).
† Correlation was significant at the alpha ≤0.05 level (2-tailed).

Table IV
Correlation of ASESq domains compared with the domains to the domains of the SPADI questionnaire

ASESq	SPADI		
	Total score	Pain	Disability
Activity (n = 92)			
Correlation	−0.70*	−0.62*	−0.71*
Significance (2-tailed)	0.000	0.000	0.000
Pain (n = 92)			
Correlation	0.45*	0.50*	0.39*
Significance (2-tailed)	0.000	0.000	0.000
Stability (n = 84)			
Correlation	0.27*	0.33*	0.22†
Significance (2-tailed)	0.012	0.002	0.045
Total (n = 92)			
Correlation	−0.70*	−0.70*	−0.66*
Significance (2-tailed)	0.000	0.000	0.000

ASESq, American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form; *SPADI*, Shoulder Pain and Disability Index.
All domains of the ASES questionnaire did have statistically significant ($P < .05$) correlations with the domains of the SPADI. The highest correlation was between the ASESq total score and SPADI Pain, whereas the weakest correlation was between ASESq stability and SPADI disability.
* Correlation was significant at the alpha ≤0.01 level (2-tailed).
† Correlation was significant at the alpha ≤0.05 level (2-tailed).

ICC of our study was substantially lower.[8,12] A possible explanation for this dissimilarity could be the longer time interval used between test and retest. The mean test-retest interval of this study was 13 (±4.4) days, whereas Celic and Goldhahn had an interval of 3-7 and 7 days, respectively.[8,12] A shorter test-retest interval makes it more likely to get a similar test-result, leading to a higher ICC, because of the carryover effects of patients, who could potentially still remember what they reported in the first questionnaire.[19] During a short interval, it is unlikely that the health status of the shoulder changes considerably, which could lead to a more consistent scoring.[19] In contrast, a long test-retest period could result in significant health change. In this study, a test-retest interval of 1-4 weeks was applied to avoid the risk of memory bias as well as change of health status, as suggested in the literature.[1,3,29,33] Piitulainen had a similar test-retest interval of 2 weeks and a comparable ICC of 0.83, supporting the hypothesis mentioned above.[22] In addition, a Bland-Altman plot was drawn for visual presentation and to provide a better insight into the variations between the test and retest compared with the ICC.[2] Although the test-retest interval varies considerably between the ASESq validation studies, the generally produced ICC is similar to this study and consistently very good, indicating an excellent test-retest reliability of the ASESq-NL.[8,20–22,35]

The domains of the Dutch ASES questionnaire quitu were compared with the corresponding domains of the validated and translated Dutch instruments SF-36 and SPADI to evaluate construct validity. A good construct validity was found for the domains of the SPADI and the SF-36 with the corresponding domains of the ASESq-NL, except for the SF-36 energy, fatigue and emotional well-being. As suggested by Berendes and Piitulainen, it is expected that these SF-36 domains do not correlate linearly with shoulder problems because of their general character.[3,22] In contrast, the ASESq-NL and SPADI had significant correlations in all domains. This was expected because of the specific character of the SPADI, which focuses on shoulder problems.

Although Celik and Piitulainen described a significant correlation for the SF-36 domain “emotional well-being,” Piitulainen suggested that this aberration could be due to “differences in, e.g. sample size, age, reason for shoulder disorder”.[8,22]

In addition, the ASESq-NL domain stability showed no significant correlation with the domain “physical function” of the SF-36 health form. Furthermore, the ASESq-NL stability section demonstrated the lowest test-retest agreement in this study. A similar finding was presented in the study of Goldhahn, who found a fair ICC and concluded that the stability section was not suitable for the evaluation of stability in total shoulder arthroplasty patients. Whether this evaluation is true and applicable to other patient groups remains unclear.

A significant correlation was found between the ASESq-NL stability domain and all domains of the shoulder-specific SPADI questionnaire. The ASESq stability section is however, as intended by the original author of the ASESq, not included in the calculation of the shoulder score index.

Strengths of this study are the good representation of the ASESq-NL target population, with an equal distribution of men and women, as well as the wide range of shoulder problems, keeping in mind that the ASESq was designed for shoulder problems.

A limitation of the study is that there is no final consensus on how to cross-culturally adapt and validate health questionnaires. Multiple guidelines are used with significant differences in their approach. As a result, the most respected guidelines were selected. These guidelines were generally also used by the other ASESq adaptation and validation studies, leading to comparable results with confirmation of reliability and construct validity of the ASESq-NL.

Conclusion

The Dutch ASES self-assessment form (Fig. 2) is a valid and reliable tool for the evaluation of shoulder problems and is permissible for implementation into the Dutch health care system.

Disclaimer

The authors, their immediate families and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Acknowledgments

The authors thank Virginia Hoel, Björn Remmerswaal and Mohamed Dinle for their contribution to the forward-backward translation process of this study.

References

1. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000;25:3186–91.
2. Berchtold A. Test–retest: agreement or reliability? Method Innov 2016;9:2059799116672875. https://doi.org/10.1177/2059799116672875.
3. Berendes T, Pilot P, Willems J, Verburg H, te Slaa R. Validation of the Dutch version of the Oxford Shoulder Score. J Shoulder Elbow Surg 2010;19:829–36. https://doi.org/10.1016/j.jse.2010.01.017.
4. Bland JM, Altman DG. Cronbach's alpha. BMJ 1997;314:572.
5. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60.
6. Brook RH, Ware JE, Davies AR, Stewart A, Donald CA, Rogers WH, et al. Conceptualization and Measurement of Health for Adults in the Health Insurance Study. Santa Monica, CA: Rand Corp; 1980.
7. Bot SDM, Terwee CB, van der Windt DAWM, Bouter LM, Dekker J, de Vet HCW. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rheum Dis 2004;63:335–41. https://doi.org/10.1136/ard.2003.007724.
8. Celik D, Atalar AC, Demirhan M, Dirican A. Translation, cultural adaptation, validity and reliability of the Turkish ASES questionnaire. Knee Surg Sports Traumatol Arthrosc 2013;21:2184–9. https://doi.org/10.1007/s00167-012-2183-3.

9. Cicchetti DV, Guidelines criteria. rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment 1994;6:284–90.
10. Cook KF, Roddey TS, Gartsman GM, Olson SL. Development and psychometric evaluation of the Flexilevel Scale of Shoulder Function. Med Care 2003;41: 823–35. https://doi.org/10.1097/01.MLR.0000068540.82233.
11. Elvers HOR, Siervelt IN, van der Heijden KWAP. Internal consistency and construct validity of the Dutch Shoulder Pain and Disability Index (SPADI Dutch Version) after subacromial decompression according to Neer. Ned Tijdschr Fysiotherap 2003;113(6):126–31.
12. Goldhahn J, Angst F, Drerup S, Pap G, Simmen BR, Mannion AF. Lessons learned during the cross-cultural adaptation of the American Shoulder and Elbow Surgeons Shoulder Form into German. J Shoulder Elbow Surg 2008;17:248–54. https://doi.org/10.1016/j.jse.2007.06.027.
13. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417–32.
14. Jackowski D, Guyatt G. A guide to health measurement. Clin Orthop Relat Res 2003;413:80–9. https://doi.org/10.1097/01.blo.0000079771.06654.13.
15. Knaut LA, Moser AD, Melo Sde A, Richards RR. Translation and cultural adaptation to the Portuguese language of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASES) for evaluation of shoulder function. Rev Bras Reumatol 2010;50:176–89.
16. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–63. https://doi.org/10.1016/j.jcm.2016.02.012.
17. Krousel-Wood MA. Practical considerations in the measurement of outcomes in healthcare. The Ochsner journal 1999;1:187–94.
18. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract 2006;12: 559–68. https://doi.org/10.1111/j.1365-2753.2006.00650.x.
19. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol 2003;56:730–5.
20. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. J Shoulder Elbow Surg 2002;11:587–94. https://doi.org/10.1067/mse.2002.127096.
21. Padua R, Padua L, Ceccarelli E, Bondi R, Alviti F, Castagna A. Italian version of ASES questionnaire for shoulder assessment: cross-cultural adaptation and validation. Musculoskelet Surg 2010;94:S85–90. https://doi.org/10.1007/s12306-010-0064-9.
22. Piitulainen K, Paloneva J, Ylinen J, Kautiainen H, Hakkinen A. Reliability and validity of the Finnish version of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section. BMC Musculoskelet Disord 2014;15:272. https://doi.org/10.1186/1471-2474-15-272.
23. Poolman RW, Swiontkowski MF, Fairbank JCT, Schemitsch EH, Sprague S, de Vet HCW. Outcome instruments: rationale for their use. J Bone Joint Surg Am 2009;91:41–9. https://doi.org/10.2106/JBJS.H.01551.
24. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg 1994;3:347–52.
25. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Care Res 1991;4:143–9.
26. Roe Y, Soberg HL, Bautz-Holter E, Ostensjo S. A systematic review of measures of shoulder pain and functioning using the International Classification of Functioning, Disability and Health (ICF). BMC Musculoskelet Disord 2013;14: 73. https://doi.org/10.1186/1471-2474-14-73.
27. Sallay PI, Reed L. The measurement of normative American Shoulder and Elbow Surgeons scores. J Shoulder Elbow Surg 2003;12:622–7. https://doi.org/10.1016/S1058-2746(03)00209-X.
28. Schmidt S, Ferrer M, González M, González N, Valderas JM, Alonso J, et al. Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. J Shoulder Elbow Surg 2014;23:434–44. https://doi.org/10.1016/j.jse.2013.09.029.
29. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury 2011;42:236–40. https://doi.org/10.1016/j.injury.2010.11.042.
30. Suk M, Hanson BP, Norvell DC, Helfet DL. AO Handbook: Musculoskeletal Outcomes Measures and Instruments. New York, NY: Thieme; 2005.
31. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011;2:53–5. https://doi.org/10.5116/ijme.4dfb.8dfd.
32. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health 2008;11:700–8. https://doi.org/10.1111/j.1524-4733.2007.00309.x.
33. van Kampen DA, van Beers LW, Scholtes VA, Terwee CB, Willems WJ. Validation of the Dutch version of the Simple Shoulder Test. J Shoulder Elbow Surg 2012;21:808–14. https://doi.org/10.1016/j.jse.2011.09.026.
34. Vrotsou K, Cuellar R, Silio F, Rodriguez MA, Garay D, Busto G, et al. Patient self-report section of the ASES questionnaire: a Spanish validation study using classical test theory and the Rasch model. Health Qual Life Outcomes 2016;14: 147. https://doi.org/10.1186/s12955-016-0552-1.
35. Yahia A, Guermazi M, Khmekhem M, Ghroubi S, Ayedi K, Elleuch MH. Translation into Arabic and validation of the ASES index in assessment of shoulder disabilities. Ann Phys Rehab Med 2011;54:59–72. https://doi.org/10.1016/j.rehab.2010.12.002.