Management of (traumatic) anterior shoulder instability: current treatment and future perspectives The open Bankart procedure still state of the art in 2020

Berendes, T.D.

3

Chapter 3

Validation of the Dutch version of the Oxford Shoulder Score

Thomas Berendes · Peter Pilot · Jaap Willems · Hennie Verburg · Ron te Slaa

J Shoulder Elbow Surg. 2010;19:829-836

Abstract

Background: The Oxford Shoulder Score (OSS) is an internationally used patient-based outcome score. Up to now, it was not validated in Dutch. The purpose of this study was to produce a Dutch translation of the OSS and to test this version in terms of reliability and validity.

Methods: Translation of the OSS was done according to the guidelines in literature. One hundred and three patients completed the Dutch version of the OSS. Additionally, the Constant-Murley shoulder score, the (Dutch) Simple Shoulder Test (DSST) score, and SF-36 were included into the validation process. Feasibility and patient-burden parameters were also tested.

Results: One-hundred and three patients with general shoulder problems age 55 years (min-max: 21-81 ± 13 yrs.), sex ratio 2/3 (f/m) completed the Dutch version of the OSS and the SF-36. Internal consistency tested by the Cronbach's alpha (0.921) was high. Intra-class correlation coefficient was R = .981 (95% confidence interval: .961–.993) and the mean difference between both tests was 2.7 points (0–8). Construct validity was also tested by the Pearson correlation coefficient and showed a significant correlation ($p < .01$) between the Dutch version of the OSS and the other scores (DSST 0.61; the Constant-Murley score 0.64 and with most of the SF-36 sub-scores, except for 2 psychometric subscales, namely, mental health (0.15 [$p = .123$]) and general health (0.10 [$p = .316$])

Conclusion: The instrument proved to be valid by demonstrating significant correlations predicted by standard clinical assessments (DSST and Constant-Murley scores) and a generic patient-based instrument (SF-36). Application and evaluation in clinical trial proved feasible and understandable.

Introduction

Outcome measures play an increasingly important role in medical practice. However, measuring these outcomes in a simple, reliable and valid way is important. Since the 1980s, a large body of research has been devoted to the development of health-related quality of life (HRQOL) measures.[17] One can make a distinction between generic instruments instead of joint-specific and disease-specific questionnaires. The generic instruments are developed for evaluation of the overall status of a patient (e.g., SF-36), and the latter is a measure that attempts to quantify function following disease, injury, or treatment of a specific joint or body part. Outcome instruments intended for patients need 2 essential requirements: 1) that it measures what it is supposed to, and 2) that this measure is made with the minimum of error, e.g., validity and reliability.[30]

To use an outcome instrument in a different language from which it was originally designed and validated, one must take into account cross-cultural differences. Cross-cultural adaptation has 2 components: the translation of the HRQOL measure and its adaptation, i.e. a combination of the literal translation of individual words and sentences from one language to another and an adaptation with regard to idiom and to cultural context and lifestyle.[17] For instance, by translation a double denial might enter, or the type of toilet used in a country might differ (French versus European toilets) giving different scores for this question in different countries. The presence of culturally equivalent instruments would allow direct international comparison of national studies.[17]

Regarding the shoulder, several patient oriented outcome measures are validated and available in Dutch. The most used are the Disabilities of the Arm, Shoulder and Hand Questionnaires (DASH), and Simple Shoulder Test (SST). The DASH is a 30-point patient-based, non-joint specific outcome measure.[21] In other words, it is a region specific outcome measure. The SST is a shoulder specific patient based outcome score consisting of 12 "yes or no" questions.[2-3,39] One of the other frequently used shoulder specific patient questionnaires is the Oxford Shoulder Score (OSS), an international widely used patient-based outcome score consisting of 12 questions.[11] However, up to now, it was not validated in Dutch. The purpose of this study was to produce a Dutch translation of the OSS and test the Dutch version in terms of reliability and validity.

Materials and methods

The study was divided into 2 phases: First, the original 12-item questionnaire was translated into Dutch. Second, this version, with the added questions, was then tested for psychometric quality in a prospective study involving 103 patients.

Translation procedure

The Questionnaire was translated by an experienced medical language editing company in the Netherlands (ISYS Prepress Services, Winkel, the Netherlands). After this process, a reverse translation was made by an English mother-tongue individual. The similarities and differences were then reviewed by a committee of 2 experienced orthopaedic shoulder surgeons, a health scientist, and a resident orthopaedic surgery. The committee debated the discrepancies, and, if needed, decided to repeat the translation-back-translation process. The complete questionnaire was then tested on 20 patients with shoulder problems in the Reinier de Graaf Gasthuis to check the comprehensibility by means of the probe technique and was then adjusted to form the final version of the Dutch Oxford Score.[17]

Prospective trial

All 103 patients were recruited into a prospective study during July 2008 to December 2008 in two departments of Orthopaedics and Traumatology in the Netherlands (RdGG / OLVG = 60 / 43). This gave us 2 groups of patients, that for this procedure (translation, cross-cultural adaptation, and validation of a measuring instrument in 2 different centers) has not been performed in other cases in the past.[11,23,29,32,43] The 103 patients suffering from degenerative and inflammatory changes, together with post-traumatic problems of the shoulder region (arthritis, cuff pathology, or tendinitis calcarean), were selected from the out-patient clinics (Table 1). All patients included were identified by 1 of the experienced shoulder surgeons (RtS, WJW). After inclusion, the patients completed the Dutch version of the OSS together with 2 other patient-based outcome questionnaires (SST, Medical Outcome Study Short Form-36). The clinician-based outcome score (Constant-Murley shoulder assessment) was completed by the orthopaedic surgeon.

Oxford Shoulder Score (OSS)

The OSS is a joint-specific patient reported outcome score, including 2 subscales, that was developed for patients with a degenerative or inflammatory state of the shoulder.[11] It is not suitable for patients with instability of the shoulder. It contains 12 items to be answered by the patient independently. There are 5 categories of response for every question, corresponding to a score ranging from 1 to 5. Scores are combined to give a single score, with a range from 12 (best) to 60 (worst). The questions deal with pain (degree, time point) and possible handicaps in private and professional life. It is divided 20/40 corresponding to pain/activities of daily living.

(Dutch) Simple Shoulder Test (DSST)

This patient-reported outcome score deals with pain and shoulder function. The (Dutch) Simple Shoulder Test (DSST) score consists of 13 simple questions of which one can choose “yes” or “no” as an answer.[39]

Table 1: Demographic data

N (patients)	103
RdGG / OLVG	60 / 43
Mean age (years ± SD)	55 (±13)
Minimum–maximum age (years)	21–81
Male–Female	52–51
Left–Right shoulder	59–44
Diagnosis	
Impingement syndrome without rotator cuff tear	35
Rotator cuff tear with/without impingement syndrome	26
Glenohumeral arthritis	12
Acromioclavicular related problems	9
Frozen shoulder	8
Cervicobrachialgy	4
Post-traumatic	9
Pseudo artrosis of clavicular fracture	3
Others (pseudo artrosis of humeral fracture; AVN after fracture of proximal humerus; malunion after fracture of proximal humerus; lateral clavicular fracture; undefined fracture, scapula lata)	6

MOS SF-36

MOS SF-36 is a generic patient reported outcome score, consisting of 36-items.[4,5,42] It is widely used to assess the general health of the patient. It provides scores on 8 dimensions or subscales: physical function, social function, limitations caused by physical symptoms, limitations caused by emotional problems, general mental health, vitality, pain, and perception of general health. Each subscale has a minimum score of 0 points and a maximum score of 100 points.

Constant-Murley shoulder assessment

The Constant-Murley functional assessment of the shoulder (1987) is a clinician based outcome score, consisting of 4 subscales (10 items).[7] The outcome of the Constant-Murley shoulder assessment score has been validated against: the OSS, the change in day-to-day life, improvement, success of operation, SF-36, DASH, ASES, and the DSST.[12,28,33]

The Dutch Oxford Score was investigated for reproducibility, internal consistency, and (construct) validity. The same set-up and statistical methods were used as with the German version of the OSS.[20] Furthermore, the results will be compared with those of the original English Oxford Shoulder Score and those of the German Oxford Shoulder score.[11,20]

Psychometric testing

Reliability

In research, the term reliability means "repeatability" or "consistency".[35] A measure is considered reliable if it would give us the same result over and over again (assuming that what we are measuring is not changing!).[36] Reliability assesses the error in an instrument. Others have referred to this as "consistency", as reliability may be confused with "trustworthy", which would not be appropriate if an instrument repeatedly yields the wrong results.[14] Like validity, reliability is not a fixed property but is dependent upon the context of the population studied.[34] However, reliability does not imply validity. That is, a reliable measure is measuring something consistently, but not necessarily what it is supposed to be measuring.[8,36]

Reproducibility

Reproducibility is a form of reliability that can be further subdivided into inter-observer and test-retest. How closely one observer agrees with another observer using the same instrument and the same patient is the essence of inter-observer reproducibility (applicable to clinician-based outcomes). Test-retest reproducibility is measured by administering the same instrument to the same patient on 2 different occasions when no important dimensions of health have changed.[35] A definition of reproducibility is: the closeness of agreement between independent results obtained with the same method on identical test material but under different conditions (different operators, different apparatus, different laboratories, and/or after different intervals of time).[26] To test reproducibility or test-retest reliability, we asked 27 of the patients included to answer the questionnaire again within 24–72 hours to see whether they completed it with the same answers. The reproducibility was investigated by calculating the intra-class correlation coefficient (ICC, 2-way random model for agreement) between the test and re-test.[25]

Internal consistency

Internal consistency is a measure based on the correlations between different items on the same test (or the same subscale on a larger test). It measures whether several items that propose to measure the same general construct produce similar scores. Internal consistency can be tested in various ways. However, Cronbach's alpha is the mostly used way.[9] High reliabilities (0.95 or higher) are not necessarily desirable, as this indicates that the items may be entirely redundant. The goal in designing a reliable instrument is for scores on similar items to be related (internally consistent), but for each to contribute some unique information as well.

Validity

Validity is an index of how well a test measures what it is supposed to measure. In this case, that meant assessing the validity of the Dutch version of the OSS. The Pearson correlation coefficient was calculated between the OSS and the Constant-Murley, DSST and SF-36. The Pearson correlation assumes that the 2 variables are measured on at least interval scales, and it determines the extent to which values of the 2 variables are "proportional" to each other. The value of correlation (i.e., correlation coefficient) does not depend on the specific measurement units used.[36]

Construct validity

There is an awful lot of confusion in the methodological literature that stems from the wide variety of labels that are used to describe the validity of measures. Any time a concept or construct is translated into a functioning and operating reality (the operationalization), there is a need to be concerned about how correct the translation is. This issue is as relevant when we are talking about treatments or programs as it is when we are talking about measures. (The population of interest in the study is the "construct" and the sample is your operationalization. If we think of it this way, we are essentially talking about the construct validity of the sampling!)[36]

Because there is no 'gold-standard', construct validity was determined by comparing the OSS with various subscales of the generic MOS SF-36, the Constant-Murley Assessment Score, and the DSST. Spearman rank correlation coefficients were calculated. To investigate whether the OSS of satisfied patients differed from the OSS of dissatisfied patients, the scores of every follow-up occasion were compared with each other. Convergent and divergent validity were measured by investigating the strength of the correlation coefficients. The OSS should converge, have high correlations, with similar metrics (VAS for pain, physical functioning) and diverge, have low correlations, from dissimilar domains from the RAND-36 (e.g., general perception of health, mental health).

Statistical evaluation

All analyses were carried out using the Statistical Package for the Social Sciences (SPSS; Gorinchem, the Netherlands), version 11.5.

Results

One-hundred and three patients completed the questionnaires and were investigated clinically (see Table 1 for the demographic data). Sixty patients were included in Delft, Reinier de Graaf Gasthuis. Forty-three patients were included in Amsterdam, Onze Lieve Vrouwe Gasthuis. The patients themselves completed all questionnaires, except

Table 2: Absolute values of the Dutch Simple Shoulder score (DSST), the Constant-Murley score (CM) for the injured and non-injured shoulder, the Dutch Oxford Shoulder score (OSS) and the MOS SF-36 with the different subscales

	DSST	CM score injured shoulder	CM score non injured shoulder	OSS	SF-36 PF	RF	BP	GH	VI	SF	RE	MH
N (patients)	103	103	103	103	103	103	103	103	103	103	103	103
Mean	6.3	60.7	86.5	32.5	69.1	41.8	44.5	66.1	64.1	72.5	48.2	74.4
Median	6.0	64	88	31	75	25	40	67	65	75	33	76
SD	3.8	20.2	10.3	9.5	21.8	39.5	22.0	19.0	18.2	24.8	42.5	17.7
Min	0	2	49	13	0	0	0	5	20	0	0	16
Max	13	95	100	55	100	100	100	100	100	100	100	100

Abbreviations: PF, physical function; RF, role physical; BP, bodily pain; GH, general health; VI, vitality; SF, social functioning; RE, role emotional; MH, mental health.

for the Constant-Murley score (which is a clinician-based outcome score). The absolute values of all scores are given in Table 2.

Reproducibility

The patients seen in Delft were asked to fill in the questionnaire twice for testing of test-retest reliability. Thirty-seven questionnaires were received. However, 10 were anonymous, leaving 27 for evaluation. The intra-class correlation coefficient was R = .981. The mean difference between both tests was 2.7 points (0–8), corrected for 1 outlier (17) (see Table 3).

Table 3: Reproducibility numbers: reliability statistics

Cronbach's Alpha	No. of Items
.981	54 (= 2 x 27)

The intraclass correlation[a] was .485[b] for the single measures and .981[c] for the average measures.
Two-way mixed effects model, where people effects are random and measures effects are fixed.
[a] Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.
[b] The estimator is the same, whether the interaction effect is present or not.
[c] This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Internal consistency

Internal consistency was high (Cronbach's alpha, 0.921). Elimination of one item in all 12 cases did not result in a value < 0.907. Only 2 items had low correlations with the total correlation (0.460 and 0.384). All other items had a correlation of > 0.644 with the total correlation score (see Table 4).

Construct validity

The construct validity was tested by the Pearson correlation coefficient (= "R") (Table 5). As assumed, there was a significant correlation between the (Dutch) OSS and the individual total scores. Only the subscales mental health and general health of the MOS SF-36 did not have a significant correlation at the 0.01 level (2-tailed; p = .123 and p = .316, respectively) (see Table 5).

Table 4: Internal consistency of the Oxford shoulder score: Item-total statistics

Question	Mean score (SD)	Corrected item-total correlation	Cronbach's Alpha if item deleted
1	3.853 (0.78)	.460	.922
2	2.961 (0.86)	.644	.916
3	3.226 (0.97)	.648	.915
4	4.069 (1.15)	.384	.927
5	2.549 (0.98)	.752	.911
6	1.784 (0.97)	.682	.914
7	1.814 (1.10)	.744	.911
8	2.245 (1.19)	.713	.913
9	2.049 (1.20)	.689	.914
10	2.667 (1.32)	.824	.907
11	2.814 (1.22)	.764	.910
12	2.423 (1.24)	.783	.909

The intraclass correlation[a] was .492[b] for the single measures and .921[c] for the average measures.
Two-way mixed effects model, where people effects are random and measures effects are fixed.
[a] Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.
[b] The estimator is the same, whether the interaction effect is present or not.
[c] This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Table 5: Correlations between (Dutch) Oxford shoulder score and Constant-Murley shoulder assessment, DSST score and MOS SF-36, compared with the original (English) and German validated versions (literature) of the OSS

	Correlation with OSS Dutch version	Correlation with OSS English version (literature)		Correlation with OSS German version (literature)	Mean correlation
		Pre ok	Post ok		
DSST score	0.61	-	-	-	-
Constant score	0.64	0.74	0.75	0.60	0.69
MOS SF-36					
Physical functioning	0.68	0.61	0.62	0.62	0.64
Role physical	0.46	0.41	0.61	0.56	0.52
Bodily pain	0.56	0.66	0.68	0.76	0.67
Vitality	0.20	0.52	0.59	0.49	0.47
Social functioning	0.25	0.55	0.61	0.45	0.48
Role emotional	0.38	0.37	0.51	0.27	0.39
Mental health	0.15 (p = 0.123)	0.39	0.54	0.54	0.41
General health	0.10 (p = 0.316)	0.34	0.42	039	0.33

Discussion

Shoulder problems are the third most frequent disorder of the locomotor system after back and neck problems.[10,27,38] Concerning these shoulder problems, there is considerable uncertainty as to the effectiveness of the various treatment methods.[10,27,38] However, we see lots of possible ways to evaluate the outcome of orthopaedic conditions: physical findings such as sensibility or strength, radiographic changes, electromyographic changes, and a variety of patient-satisfaction and functional criteria are all used. This is often well-defined and standardized.[1,24] Moreover, there is a demand for instruments enabling comparison of treatment results and, thereby, international studies. Nowadays, there are lots of different questionnaires, all with their own advantages and disadvantages.[35] There are, of course, comparative studies to evaluate and criticize between the different outcome measurement instruments. However, when comparing earlier performed studies, one is sometimes committed to 1 or 2 specific questionnaires. This gives an increasing need for outcome measurements overall, but also validated for the Dutch speaking countries. Therefore, most used questionnaires ought to be validated and available in different languages. Outcome scores are originally developed in specific languages and when translated must be validated in the native language of the patient before use. In the Anglo-Saxon area, several shoulder measurements instruments, like the Constant-Murley, UCLA rating scale, and Rowe scores, are available. Besides these clinician-based outcome scores, several patient based outcome scores are also available, such as DASH (quick-DASH), WOSI, Oxford instability score, and the OSS. All of these were tested psychometrically. All of the above-mentioned measurement instruments have their own differences, such as structure, size, and indication.

When looking at the concerns and priorities of patients and surgeons, we often see differences. Therefore, it is increasingly recognized that methods are required to elicit the patient's perception of the outcome.[44] Furthermore, patient-based outcome measures are less time consuming for the clinician. This is an important factor in treatment of patients. There are also advantages in using questionnaires designed to address the patient's perception of a single condition. These usually shorter questionnaires may be just as sensitive to changes of importance to patients and much simpler to use.[15,22] Another advantage of patient-based outcome scores is that they are easy to use in daily practice and are cheap applications for general use.

Then again, in 2003, Ragab showed us a source of discrepancy between self-administered patient outcome questionnaires and the outcome measures developed and administered by clinicians.[31] He also found that patients' expectations had changed from their pre-operative expectations. Although outcome measures developed and administered by clinicians are subject to bias from several sources, results of this study suggested that self-administered patient outcome measures also have their disadvantages. The validity of self-administered patient outcome questionnaires can be severely impacted by the

patients' understanding of the questions asked, as even the most seemingly simple questions are subject to misinterpretation. In our study, we tried to solve that problem by testing the understanding of the Dutch version of the OSS by means of the probe technique for the first 20 patients in the study.[17] The OSS has a 5-point Likert system specifically developed for the subjective evaluation of patients with degenerative or inflammatory changes of the shoulder, which enables quick answering by the patients as well as uncomplicated and time-saving evaluation by the investigator, offering a distinct advantage for clinical routine.[18,20] For the Dutch speaking area, there already exists translations of the Disabilities of the Arm, Shoulder and Hand (DASH), and the DSST.[39,40] But until now, the OSS has not been validated in Dutch and the OSS was chosen for the same reasons as Huber et al in his 2004 German translation.[20] The structure of the questions of the OSS is simple and easily understood resulting in an easy acceptance for patients (quick filling-in and time saving for the examiner and easy administration, especially when using computerized patient self-assessment software, with touchscreen possibilities).[6]

Reproducibility, or the test-retest reliability, was performed with 27 of the 103 patients (Table 3). When starting the study, the intentions were to use all the patients in Delft to reproduce a second form. The Delft patients were to answer the questionnaire again within 24-48 hours to see whether they completed it with the same answers. However, we ended up with 27 subjects for this sub population. Looking at the sub-group of responders (N = 27) compared to the total group (N = 76) after 24-48 hours, we saw no major differences in the outcome of the OSS (respectively, mean = 31.4 ± 8.6; min-max = 17–49 vs mean = 32.8 ± 9.9; min-max 13–55). It is not likely that their shoulder problems had changed in this short interim period. This was shown by the intra-class correlation coefficient of the test-retest population, R = .981. This is a very high outcome, nearly 1.0, meaning that almost all 27 answered their questionnaire exactly the same way at 2 different points of times.

When looking at the results of our study, we saw a high internal consistency (Table 4). Examination of reliability resulted in a Cronbach's coefficient alpha of R = .921, an excellent value. Elimination of 1 item in all 12 cases did not result in a value < 0.907. Questions 1 and 4 showed the lowest corrected item-total correlation scores, namely 0.460 and 0.384, respectively. This suggests that those 2 items are not closely related (internally consistent) to the rest of the OSS, but each gives their own unique information. If we look at the questions closely (see Appendix), we see that for all questions, except questions 1 and 4, it is a reflection of daily live activities. When looking precisely, question 1 is presented in the past tense related to general pain problems of the shoulder. It tries to give a reflection of the pain someone had experienced at any time in the past (not necessarily recently) regarding the shoulder. Question 4 also concerns pain problems of the shoulder, but at rest when lying in bed at night (different from

daily live activities). On the other hand, as said earlier, a very high Cronbach's coefficient alpha score of almost 1.0 is not necessarily good, because then the question becomes redundant. It will not give us new information.

In this study, we saw a fair to moderate correlation of the Dutch version of the OSS against the DSST, MOS SF-36, and Constant-Murley score. However, this is not unexpected, as these outcome assessments are not identical. Subsequently, one could have compared the (Dutch) OSS with the more likely assessment of the (Dutch) DASH. However, this was not done. The construct validity (Table 5), presented here with the correlation coefficients between the absolute values of the (Dutch) OSS, DSST, Constant-Murley, and the relevant sub-scores of the MOS SF-36, was generally high (R > .46). Again, the subscales for pain and physical functioning of the SF-36 score exhibited the highest values. In our study, again we noticed that the shoulder questionnaires DSST and Constant-Murley score performed substantial higher correlation rates for the Dutch OSS compared with most of the different non-physical subscales of the SF-36. This confirms the need to use both joint-specific and generic health-status measures to evaluate patients who have a problem related to the shoulder. Noting our Materials and Methods section, we can see from our results that indeed the (Dutch) OSS has high correlations with similar metrics (e.g., VAS for pain, DSST and Constant-Murley score) and lower correlations with dissimilar domains from the MOS SF-36 (e.g., general perception of health, mental health). The explanation for this difference seems logical due to the direct relation between function or dysfunction of the shoulder ("physical functioning", "role physical", "bodily pain" of a patient) and joint specific shoulder scores. It is not necessarily true that if someone has shoulder problems that he or she also has a diminished vitality or social functioning (same with "role emotional", "mental health" and "general health").

In 1996, Dawson et al came with the source of the OSS in which the patient population was tested in patients with chronic shoulder complaints.[11] In 2001, the OSS was tested in patients who underwent rotator cuff surgery.[12] More recently, Dawson et al published the OSS revisited.[13] Othman and Taylor tested the OSS in patients with frozen shoulders.[28] So, the OSS was tested in most type of shoulder complaints. Afterwards, the OSS was also validated in other languages. In 2004, Huber et al presented his German version of the OSS.[20] We validated the Dutch version of the OSS based on this study of Huber et al.[20]

As previously mentioned, it is necessary to validate a questionnaire into the native language. For the shoulder specific questionnaires, this was done for the Shoulder Disability Questionnaire, the DASH, DSST, and the Shoulder Rating Questionnaire (SRQ).[37,39-41] In Oxford scores for other parts of the body, this process of validation in Dutch was done for the Oxford Hip and the Oxford knee scores.[16,19]

Comparing our results with those of the original English Oxford Shoulder Score and those of the German Oxford Shoulder score, this study shows that the correlations between (Dutch) Oxford shoulder score and Constant-Murley shoulder assessment and DSST score and the physical health subscales of the MOS SF-36 are more or less comparable with the original (English) and German validated versions (literature) of the OSS. Again, the subscales for pain and physical functioning of the SF-36 exhibited the highest values, and thus corresponded with the values published for the English and German versions. An odd thing was that the Dutch correlations of the mental health subscales of the MOS SF-36 (vitality, social functioning, role emotional, mental health and general health) were somewhat lower compared with the values published for the English and German versions. Especially the subscales mental health (DOSS: 0.15 vs EOSS: 0.39 pre-op vs GOSS: 0.54) and general health (DOSS: 0.10 vs EOSS: 0.34 vs GOSS: 0.39) turned out to be substantially lower. The explanation for this is not clear. For the German study group, the demographic data was at least different from the Dutch and English groups for age characteristics (mean ages; DOSS: 55 ± 13 yrs. vs EOSS: 57 ± 15 yrs. vs GOSS: 54 ± 10 yrs.). But the Dutch and English groups were comparable for age. Another explanation might be that our study population consisted of 2 different sub-populations: rural (Delft) and urban (Amsterdam). However, these subgroups did not differ for the OSS (respectively, mean = 32.7 ± 9.4; min-max = 17–55 vs mean = 32.1 ± 9.8; min-max 13–55). A possible point of criticism is that we did not perform a power analysis when starting the study, but used the number reported by Huber et al.[20] However, looking at the post-hoc performed power analysis, we see that the decision error (= ß) is 0.1 when the sample size exceeds 70, assuming that our zero-hypothesis is that there is no difference between the different tests (for Confidence Interval [= C.I.] is 95% and effect size [= d] is 5%). Our sample size was 103. Furthermore, the composition of our study was a heterogeneous population if compared to the German population. One could debate whether this is a weakness or strength of the validation process. Finally, we did not include sensitivity to change in our study. "Sensitivity to change" or responsiveness is the ability of a measure to detect a change when a change has occurred, in particular, changes in response to some intervention. This would have added strength to the validation process.

Conclusion

We validated and tested the OSS short 12-item questionnaire in Dutch, which patients found easy to complete and proved to provide reliable, valid, and responsive data regarding their perception of general shoulder problems (excluding instability problems). Our results showed that the Dutch version of the OSS, which is intended for use as an outcome measure during specialist treatment, imposes very little burden on the patients.

References

1. Amadio PC. Outcomes measurements. J Bone Joint Surg Am 1993;75: 1583-4.
2. Beaton D, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. J Shoulder Elbow Surg 1998;7:565-72.
3. Beaton DE, Richards RR. Measuring function of the shoulder. A cross-sectional comparison of five questionnaires. J Bone Joint Surg Am 1996;78:882-90.
4. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992; 305:160-4.
5. Bullinger M. German translation and psychometric testing of the SF-36 Health Survey: preliminary results from the IQOLA Project. International Quality of Life Assessment. Soc Sci Med 1995; 41:1359-66.
6. Chinman M, Hassell J, Magnabosco J, Nowlin-Finch N, Marusak S, Young AS. The feasibility of computerized patient self-assessment at mental health clinics. Adm Policy Ment Health 2007; 34:401-9.
7. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop Relat Res 1987; 214:160-4.
8. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol 1993; 78:98-104.
9. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrica 1951:297-334.
10. Dalton SE. The conservative management of rotator cuff disorders. Br J Rheumatol 1994; 33:663-7.
11. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br 1996;78:593-600.
12. Dawson J, Hill G, Fitzpatrick R, Carr A. The benefits of using patient-based methods of assessment. Medium-term results of an observational study of shoulder surgery. J Bone Joint Surg Br 2001; 83:877-82.
13. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford shoulder score revisited. Arch Orthop Trauma Surg 2009; 129:119-23.
14. Feinstein AR. Clinimetrics. Yale University Press, New Haven 1987.
15. Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A. A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. J Rheumatol 1993; 20:429-36.
16. Gosens T, Hoefnagels NH, de Vet RC, Dhert WJ, van Langelaan EJ, Bulstra SK, et al. The "Oxford Heup Score": the translation and validation of a questionnaire into Dutch to evaluate the results of total hip arthroplasty. Acta Orthop 2005; 76:204-11.
17. 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.
18. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. J Chronic Dis 1987; 40:1129-33.
19. Haverkamp D, Breugem SJ, Sierevelt IN, Blankevoort L, van Dijk CN. Translation and validation of the Dutch version of the Oxford 12-item knee questionnaire for knee arthroplasty. Acta Orthop 2005; 76:347-52.
20. Huber W, Hofstaetter JG, Hanslik-Schnabel B, Posch M, Wurnig C. The German version of the Oxford Shoulder Score: cross-cultural adaptation and validation. Arch Orthop Trauma Surg 2004; 124:531-6.
21. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). Am J Ind Med 1996; 29:602-8.

22. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. Med Care 1992; 30:917-25.
23. Kessler S, Pfander T, Nelitz M, Puhl W, Gunther KP. The Pediatric Musculoskeletal Functional Health Questionnaire. A function assessment questionnaire for detection of illnesses of the locomotor system in children and adolescents: initial results of validating a German version. Z Orthop Ihre Grenzgeb 2001; 139:134-7.
24. MacDonald DA. The shoulder and elbow. In: Pynsent PB, Fairbank JC, Carr A, editors. Outcome measures in orthopaedics. Oxford: Butterworth Heinemann; 1993. p. 144-73.
25. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996; 1:30-46.
26. McNaught AD. IUPAC compendium of chemical terminology.2nd. United Kingdom: Blackwell Science; 1997.
27. Ogilvie-Harris DJ, Demaziere A. Arthroscopic debridement versus open repair for rotator cuff tears. A prospective cohort study. J Bone Joint Surg Br 1993; 75:416-20.
28. Othman A, Taylor G. Is the constant score reliable in assessing patients with frozen shoulder? 60 shoulders scored 3 years after manipulation under anaesthesia. Acta Orthop Scand 2004; 75:114-6.
29. Pose B, Sangha O, Peters A, Wildner M. Validation of the North Amer-ican Spine Society Instrument for assessment of health status in patients with chronic backache. Z Orthop Ihre Grenzgeb 1999; 137:437-41.
30. Pynsent PB, Carr AJ. Outcome Measures in Orthopaedics and Orthopaedic Trauma.2nd. Great Britain: Arnold; 2004.
31. Ragab AA. Validity of self-assessment outcome questionnaires: patient-physician discrepancy in outcome interpretation. Biomed Sci Instrum 2003; 39:579-84.
32. Rosales RS, Delgado EB. Diez de la Lastra-Bosch I. Evaluation of the Spanish version of the DASH and carpal tunnel syndrome health-related quality-of-life instruments: cross-cultural adaptation process and reliability. J Hand Surg [Am] 2002; 27:334-43.
33. Skutek M, Fremerey RW, Zeichen J, Bosch U. Outcome analysis following open rotator cuff repair. Early effectiveness validated using four different shoulder assessment scales. Arch Orthop Trauma Surg 2000; 120:432-6.
34. Streiner DL. Health measurement scales: a practical guide to their development and use. Oxford, UK: Oxford University Press; 1995.
35. Suk M, Hanson BP, Norvell DC, Helfet DL. Musculoskeletal Outcomes Measures and Instruments. AO Handbook. Switzerland: AO publishing; 2005.
36. Trochim WM, Donnelly JP. The Research Methods Knowledge Base. Mason, OH: Thomson Publishing; 2007.
37. van der Heijden GJ, Leffers P, Bouter LM. Shoulder disability questionnaire design and responsiveness of a functional status measure. J Clin Epidemiol 2000; 53:29-38.
38. van der Windt DA, van der Heijden GJ, Scholten RJ, Koes BW, Bouter LM. The efficacy of non-steroidal anti-inflammatory drugs (NSAIDS) for shoulder complaints. A systematic review. J Clin Epidemiol 1995; 48:691-704.
39. van Laarhoven HA, Te Slaa R, De Boer LM, Willems WJ. De Nederlandse Simpele Schouder Test, een efficiente wijze van schouderfunctiebepaling. NTvO 2001; Vol. 8.
40. Veehof MM, Sleegers EJ, van Veldhoven NH, Schuurman AH, van Meeteren NL. Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). J Hand Therapy 2002; 15:347-54.

41. Vermeulen HM, Boonman DC, Schuller HM, Obermann WR, van Houwelingen HC, Rozing PM, et al. Translation, adaptation and validation of the Shoulder Rating Questionnaire (SRQ) into the Dutch language. Clin Rehabil 2005; 19:300-11.
42. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I Conceptual framework and item selection. Med Care 1992; 30:473-83.
43. Wiesinger GF, Nuhr M, Quittan M, Ebenbichler G, Wolfl G, Fialka-Moser V. Cross-cultural adaptation of the Roland-Morris questionnaire for German-speaking patients with low back pain. Spine 1999;24: 1099-103.
44. Wright JG, Rudicel S, Feinstein AR. Ask patients what they want. Evaluation of individual complaints before total hip replacement. J Bone Joint Surg Br 1994; 76:229-34.

APPENDIX

Discussion on instability specific measurement tools with attention to the Oxford Shoulder Score

Patient perception on shoulder instability (PROM's)

A well-designed study that clearly delineates superiority of one treatment over another may provide insufficient evidence or even be harmful if it fails to measure "important" outcomes. With the pressure to approve and recommend specific devices and interventions to the musculoskeletal clinical and research community, musculoskeletal outcome research has a unique opportunity to demonstrate its ability to validate the true clinical benefit of these modalities in appropriate patient populations using appropriate measures and instruments as endpoints.[1–3]

An increasing number of outcome measurement tools have been designed to report on the effectiveness of treatment for shoulder pathologies (shoulder specific questionnaires). This includes objective measures such as range of motion (ROM), strength testing, physical exam maneuvers, return to play, and complications such as redislocation, as well as subjective measures such as patient satisfaction and patient-reported outcome (PRO) tools. However, nowhere are there more insensitive, unreliable, unvalidated measurement tools than in the orthopaedic literature.[4] These metrics are being variably applied to (shoulder) patients in literature, with a lack of consensus regarding the most appropriate method of outcome assessment for patients with (surgically managed) anterior glenohumeral instability (condition-specific questionnaires).[5] For that matter, most of nowadays commonly used scoring systems show major deficiencies when applied to instability populations.[6,7] Yet, shoulder instability is the most common diagnosis in which condition-specific measurement tools are being used. The presentation of patients with symptomatic instability is actually different compared with other shoulder pathology. Normally, they do not present themselves having pain or a decreased function as is more common with other shoulder diagnoses. This leads to poor responsiveness and significant ceiling effects when general shoulder-specific measures are being used for patients with instability problems. Therefore, more condition-specific instability measurement tools have been developed to evaluate patients with shoulder instability aiming for more response to treatment effect.

The **Oxford Shoulder Score** (OSS) is being developed by Dawson et al. in 1996.[8,9] Two different questionnaires exist. One was originally constructed for shoulder operations other than instability (OSS) in 1996. This one was translated in Dutch in 2010 (*see **attachment A** for full Dutch version of questionnaire*).[10] The other was constructed in 1999 specifically for assessment of shoulder instability patients, now called the Oxford Shoulder Instability Score (OSIS) or Shoulder Instability Questionnaire. The latter being translated in Dutch in 2015 (*see **attachment B** for full Dutch version of questionnaire*).[11]

The OSS and OSIS are both patient-based, shoulder-specific scoring systems consisting of 12 questions each with 5-graded options. The total score is given as the sum of the 12 responses (min–max: 12–60 points). The OSS questionnaire deals with the perceptions of patients about shoulder surgery and assesses pain and activities of daily living (33.3% and 66.6% respectively). It was initially based on a sample of 111 patients undergoing shoulder surgery, but excluded patients with instability. Later on, the OSS was being tested on patients undergoing rotator cuff surgery and on patients with persistent stiffness of their shoulder undergoing mobilization of their shoulder under general anesthesia. The OSIS questionnaire is a self-reported outcome measurement more specific for patients with shoulder instability. It is based on a prospective study of 92 patients with shoulder instability with both the Oxford instability and Rowe scores showing excellent responsiveness in this cohort.[8,9] More recently, study outcome also supports the use of the OSIS in military patients undergoing shoulder stabilisation surgery, probably having increased demands of military service compared to civilian patients.[12]

When looking in details of both oxford shoulder questionnaires, only two out of twelve questions of the OSIS go specific into content of shoulder instability (namely questions 1 & 5 of the OSIS, *see attachments A and B for full Dutch versions of both questionnaires*). The rest of the OSIS questions and all of the OSS questions can be interpreted on all shoulder specific conditions (other than instability). While the primary complaint in the patient with shoulder instability, and sometimes the only complaint, is apprehension or avoidance of activity. So, one can debate about an insufficient number of items in the evaluation for the instability problem in the OSIS, creating a possible deficiency in both outcome tools for that matter. Almost all other questions are about things like pain and possible limitations in the personal and professional life of the patient, on account of problems on behalf of their shoulder. Where the original purpose of the OSIS measurement tool should be to evaluate subjects with an injury-specific condition, here (in-)stability of the shoulder.

Other shoulder specific scoring systems, currently being used in literature with the intention to address shoulder instability, in no particular order, include the American Shoulder and Elbow Surgeons (ASES) score;[13,14] the Melbourne Instability Shoulder Scale (MISS);[15] the Constant-Murley (CM) score;[16] the Athletic Shoulder Outcome Rating Scale (ASORS);[17] the University of California, Los Angeles (UCLA) score;[18] the Subjective Shoulder Rating System (SSRS);[19] the L'Ìnsalata shoulder rating system / Shoulder Rating Questionnaire (SRQ);[20,21] the Disabilities of the Arm, Shoulder and Hand (DASH) score;[22,23] the Western Ontario Shoulder Instability Index (WOSI);[24,25] and the Simple Shoulder Test (SST).[26,27] The latter four are, as well as the OSS and OSIS also translated and validated for the Dutch language. However, of all the above mentioned shoulder specific questionnaires, many are actually not condition-specific for shoulder instability. See Table 1 for an overview of these shoulder-specific shoulder scoring systems, including % of instability content.

Table 1: Overview of scoring systems that are being used in literature addressing shoulder instability

Outcome measure tool	% of instability content	Year of Dutch validation	No. of questions	Total score	S/O
Rowe / modified Rowe (rating sheet for Bankart repair)	50 of total 100 points	NA	3	100	S/O
American Shoulder and Elbow Surgeon (ASES) score	Instability (and impingement) do not contribute to the total score	NA	11	100	S
L'insalata shoulder rating system (Shoulder Rating Questionnaire (SRQ))	Originally not designed for instability	2005[20]	21	100	S
Melbourne Instability Shoulder score (MISS)	33 points	NA	24	100	S
Disabilities of the Arm, shoulder and Hand (DASH) score	Originally not designed for instability	2019[23]	30	100	S
Western Ontario Shoulder Instability Index (WOSI)	100 points; 1 specific instability question (subscale A Q8)	2014[24]	21	2100	S
Oxford Instability Score (OIS) / Shoulder Instability Questionnaire	10 of 60 points	2015[32]	12	60	S
Constant-Murley (CM) score	Not appropriate for instability	NA	7	100	S/O
Athletic Shoulder Outcome Rating Scale (ASORS)	Stability 10 points	NA	6	100	S/O
University of California, Los Angeles (UCLA) score	Originally not designed for instability	NA	5	35	S/O
Simple Shoulder test (STT)	Originally not designed for instability	2001[26]; 2012[27]	12	12	S
Subjective Shoulder Rating System (SSRS)	15 points	NA		100	

Abbreviations: NA, not applicable; S, subjective; O, objective.

Nowadays, the most common validated patient reported outcome measures for shoulder instability are the Western Ontario Shoulder Instability Index (WOSI), the Melbourne Instability Shoulder Scale (MISS) and the Oxford Shoulder Instability Score (OSIS).[9,15,25,28] However, the *most commonly used* score for the evaluation of instability, is the Rowe score, which was also the first shoulder score described in 1978.[29] The Rowe score, similar to the UCLA shoulder score, was described before modern psychometric development was implemented limiting its psychometric properties.[29] The WOSI, MISS and OSIS have been developed with recent psychometric evaluations. The properties of

these scores are being described in Table 2. The WOSI is more responsive to treatment of instability than the Rowe score in patients both non-operatively and operatively treated for traumatic instability. Overall, the WOSI has the strongest psychometric properties and has undergone the most rigorous testing despite the fact that the Rowe is the most commonly reported instability measure.[30] Plancher et al. recommends in his "analysis of evidence-based medicine for shoulder instability" study, to use both the WOSI and the MISS to evaluate patients with instability for reasons of clinical-decision making.[7] However, Sahinoglu suggested specific for Dutch users to use the WOSI and the OISS for the assessment of shoulder instability patients, probably due to the fact that the MISS is still not translated into Dutch until today.[31]

Table 2: Properties of the WOSI, OSIS, MISS and Rowe scores

Instability	Description	Validity	Reliability	Responsiveness	MCID
WOSI	21 items: Physical symptoms (10) Sport / recreation / work function (4) Lifestyle function (4) Emotional function (3); Score: 0–2100 (Lower = Better) (can be converted into 0%–100% scale)	Content validity: Items established by experts and patients Criterion validity: Excellent Correlate: VAS Function and DASH, good with CMS and Rowe	Excellent ICC: 0.87–0.98	Excellent effect size: 1.67 for stabilization	220 / 2100
OSIS	12 items: Score: 12–60 (Lower = Better)	Criterion validity: Correlated with Rowe and Constant scores	Excellent ICC: 0.97	Very good effect size: 0.8	Not reported
MIIS	22 items: Pain (4) Instability (5) Function (8) Occupation and sports (5) Score: 0–100 (lower = better)	Criterion validity: Low to moderate correlation with shoulder rating questionnaire. Otherwise untested	Excellent ICC: 0.98	Not reported	Not reported
Rowe score	3 items: Stability (50 points) Motion (20 points) Function (30 points) Score: 0–100 (both subjective and examination dependent)	Content Validity: poorly described development and methodology Criterion Validity: Correlated with WOSI and CMS	Fair ICC: 0.7	Very good effect size: 1.2	Not reported

Abbreviation: ICC, intraclass correlation coefficient.

References

1. Clancy CM, Eisenberg JM. HEALTH CARE POLICY:Outcomes Research: Measuring the End Results of Health Care. *Science (80-)*. 1998;282(5387):245-246. doi:10.1126/science.282.5387.245
2. Simmons BP, Swiontkowski MF, Evans RW, Amadio PC, Cats-Baril W. Outcomes assessment in the information age: available instruments, data collection, and utilization of data. *Instr Course Lect*. 1999;48:667-685. http://www.ncbi.nlm.nih.gov/pubmed/10098096. Accessed July 12, 2019.
3. Anderson C. Measuring what works in health care. *Science (80-)*. 1994;263(5150):1080-1080. doi:10.1126/science.8108722
4. Kirkley A, Griffin S. Development of disease-specific quality of life measurement tools. *Arthrosc J Arthrosc Relat Surg*. 2003;19(10):1121-1128. doi:10.1016/j.arthro.2003.10.028
5. Lukenchuk J, Sims LA, Shin JJ. Variability in Outcome Reporting for Operatively Managed Anterior Glenohumeral Instability: A Systematic Review. *Arthroscopy*. 2017;33(2):477-483. doi:10.1016/j.arthro.2016.07.027
6. Gerber C. Integrated scoring systems for the functional assessment of the shoulder. In: *The Shoulder: A Balance of Mobility and Stability.* Rosemont Il; 1992:531-550.
7. Plancher KD, Lipnick SL. Analysis of Evidence-Based Medicine for Shoulder Instability. *Arthrosc J Arthrosc Relat Surg*. 2009;25(8):897-908. doi:10.1016/j.arthro.2009.03.017
8. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br*. 1996;78(4):593-600. http://www.ncbi.nlm.nih.gov/pubmed/8682827. Accessed September 30, 2018.
9. Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability. The development and validation of a questionnaire. *J Bone Joint Surg Br*. 1999;81(3):420-426. http://www.ncbi.nlm.nih.gov/pubmed/10872358. Accessed March 5, 2019.
10. Berendes T, Pilot P, Willems J, Verburg H, te Slaa R. Validation of the Dutch version of the Oxford Shoulder Score. *J Shoulder Elb Surg*. 2010;19(6). doi:10.1016/j.jse.2010.01.017
11. van der Linde JA, van Kampen DA, van Beers LWAH, van Deurzen DFP, Terwee CB, Willems WJ. The Oxford Shoulder Instability Score; validation in Dutch and first-time assessment of its smallest detectable change. *J Orthop Surg Res*. 2015;10(1):146. doi:10.1186/s13018-015- 0286-5
12. Middleton S, Guyver P, Boyd M, Brinsden M. Patient reported outcome measures in military patients with shoulder instability. *J R Army Med Corps*. 2013;159(4):271-273. doi:10.1136/jramc-2013-000050
13. Richards RR, An KN, Bigliani LU, et al. A standardized method for the assessment of shoulder function. *J shoulder Elb Surg*. 1994;3(6):347-352. doi:10.1016/S1058-2746(09)80019-0
14. Skutek M, Fremerey RW, Zeichen J BU. Outcome analysis following open rotator cuff repair. Early effectiveness validated using four different shoulder assessment scales. - PubMed - NCBI. *Arch Orthop Trauma Surg* . 2000;120(7-8):432-436. https://www.ncbi.nlm.nih.gov/pubmed/10968533. Accessed May 19, 2019.
15. Watson L, Story I, Dalziel R, Hoy G, Shimmin A, Woods D. A new clinical outcome measure of glenohumeral joint instability: The MISS questionnaire. *J Shoulder Elb Surg*. 2005;14(1):22-30. doi:10.1016/j.jse.2004.05.002
16. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res*. 1987;(214):160-164. http://www.ncbi.nlm.nih.gov/pubmed/3791738.
17. Tibone JE. Evaluation of treatment outcomes for the athlete`s shoulder. In: *The Shoulder: A Balance of Mobility and Stability*. 1993.
18. Amstutz HC, Sew Hoy AL, Clarke IC. UCLA anatomic total shoulder arthroplasty. *Clin Orthop Relat Res*. (155):7-20. http://www.ncbi.nlm.nih.gov/pubmed/7226634. Accessed May 19, 2019.

19. Kohn D, Geyer M. The subjective shoulder rating system. *Arch Orthop Trauma Surg*. 1997;116(6-7):324-328. doi:10.1007/BF00433982
20. L'Insalata JC, Warren RF, Cohen SB, Altchek DW, Peterson MG. A self-administered questionnaire for assessment of symptoms and function of the shoulder. *J Bone Joint Surg Am*. 1997;79(5):738-748. http://www.ncbi.nlm.nih.gov/pubmed/9160947. Accessed May 19, 2019.
21. Vermeulen HM, Boonman DC, Schüller HM, et al. Translation, adaptation and validation of the Shoulder Rating Questionnaire (SRQ) into the Dutch language. *Clin Rehabil*. 2005;19(3):300- 311. doi:10.1191/0269215505cr811oa
22. Hudak PL, Amadio PC, Bombardier C, et al. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med*. 1996;29(6):602-608. doi:10.1002/(SICI)1097-0274(199606)29:6<602::AID-AJIM4>3.0.CO;2-L
23. Veehof MM, Sleegers EJA, van Veldhoven NHMJ, Schuurman AH, van Meeteren NLU. Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). *J Hand Ther*. 15(4):347-354. http://www.ncbi.nlm.nih.gov/pubmed/12449349. Accessed March 5, 2019.
24. van der Linde JA, Jaap Willems W, van Kampen DA, A H van Beers LW, van Deurzen DF, Terwee CB. *Measurement Properties of the Western Ontario Shoulder Instability Index in Dutch Patients with Shoulder Instability*. Vol 15.; 2014. doi:10.1186/1471-2474-15-211
25. Kirkley A, Griffin S, McLintock H, Ng L. The Development and Evaluation of a Disease-Specific Quality of Life Measurement Tool for Shoulder Instability. *Am J Sports Med*. 1998;26(6):764- 772. doi:10.1177/03635465980260060501
26. van Laarhoven HA, Te Slaa R, De Boer LM, Willems WJ. De Nederlandse Simpele Schouder Test, een efficiente wijze van schouderfunctie bepaling. *Ned Tijdschr voor Orthop*. 2001;8.
27. van Kampen DA, van Beers LWAH, Scholtes VAB, Terwee CB, Willems WJ. Validation of the Dutch version of the Simple Shoulder Test. *J Shoulder Elb Surg*. 2012;21(6):808-814. doi:10.1016/j.jse.2011.09.026
28. van der Linde JA, van Kampen DA, van Beers LWAH, van Deurzen DFP, Terwee CB, Willems WJ. The Oxford Shoulder Instability Score; validation in Dutch and first-time assessment of its smallest detectable change. *J Orthop Surg Res*. 2015;10(1):146. doi:10.1186/s13018-015- 0286-5
29. Rowe CR, Patel D, Southmayd WW. The Bankart procedure: a long-term end-result study. *J Bone Jt Surg Am*. 1978;60(1):1-16.
30. Wylie JD, Beckmann JT, Granger E, Tashjian RZ. Functional outcomes assessment in shoulder surgery. *World J Orthop*. 2014;5(5):623. doi:10.5312/wjo.v5.i5.623
31. Şahinoğlu E, Ergin G, Ünver B. Psychometric properties of patient-reported outcome questionnaires for patients with musculoskeletal disorders of the shoulder. *Knee Surgery, Sport Traumatol Arthrosc*. February 2019:1-15. doi:10.1007/s00167-019-05369-7
32. Van der Linde JA, van Kampen DA, van Beers LWAH, van Deurzen DFP, Terwee CB, Willems WJ. The Oxford Shoulder Instability Score; validation in Dutch and first-time assessment of its smallest detectable change. *J Orthop Surg Res*. 2015;10(1):146. doi:10.1186/s13018-015- 0286-5

Full Dutch versions of Oxford shoulder (A) and Oxford shoulder instability score (B)

A. Oxford Schouder score (Nederlandse versie, versie 2010)[8,10]

Datum: _ _/ _ _/ 20_ _

Iedereen die deze vragenlijst gebruikt dient dit te doen onder de conditie, dat het gaat om de ***situatie van de afgelopen 4 weken***.

De test bestaat uit 12 vragen die gaan over pijn en mogelijke beperkingen in het persoonlijke en professionele leven van de patiënt.

Een hoge score (hoogste score is 60) betekent veel pijn of beperking in het functioneren ten gevolge van schouderklachten en een lage score (laagste score is 12) betekent weinig pijn of beperking in het functioneren ten gevolge van schouderklachten.

Aangedane zijde		Dominante zijde	

1. Hoe zou u de ergste pijn, die u in uw schouder heeft gehad, willen beschrijven?

Ondraaglijk	Erg pijnlijk	Nogal pijnlijk	Beetje pijnlijk	Helemaal niet pijnlijk

2. Hoe zou u de pijn, die u meestal in uw schouder heeft, willen beschrijven?

Ondraaglijk	Erg pijnlijk	Nogal pijnlijk	Beetje pijnlijk	Helemaal niet pijnlijk

3. Hoeveel beïnvloedt de pijn aan de schouder uw dagelijkse werkzaamheden? (ook het dagelijks huiswerk).

Totaal	Grotendeels	Matig	Klein beetje	Geheel niet

4. Hebt u 's nachts als u in bed ligt pijn in de schouder?

Elke nacht	De meeste nachten	Sommige nachten	1 of 2 nachten	Nooit

5. Bent u in staat u aan- en uit te kleden met uw aangedane arm?

Nee, onmogelijk	Heel erg beperkt	Matig beperkt	Nagenoeg niet beperkt	Geen beperking

6. Bent u in staat in - en uit een auto te stappen, of gebruik te maken van het openbaar vervoer met uw aangedane arm?

Nee, onmogelijk	Heel erg beperkt	Matig beperkt	Nagenoeg niet beperkt	Geen beperking

7. Kunt u op hetzelfde moment mes en vork gebruiken?

Nee, onmogelijk	Met veel moeite	Enigszins moeilijk	Zonder veel moeite	Ja, eenvoudig

8. Kunt u de boodschappen voor het huishouden zelfstandig doen?

Nee, onmogelijk	Met veel moeite	Enigszins moeilijk	Zonder veel moeite	Ja, eenvoudig

9. Kon (kunt u) u een dienblad met daarop een bord eten door de kamer dragen?

Nee, onmogelijk	Met extreem veel problemen	Met nogal wat problemen	Met lichte problemen	Ja, eenvoudig

10. Kunt u met de aangedane arm uw haar borstelen of kammen?

Nee, onmogelijk	Met veel moeite	Enigszins moeilijk	Weinig moeite	Ja, eenvoudig

11. Kunt u uw kleding in de kledingkast hangen met de aangedane arm?

Nee, onmogelijk	Met veel moeite	Enigszins moeilijk	Weinig moeite	Ja, eenvoudig

12. Kunt u zichzelf onder beide oksels wassen en drogen?

Nee, onmogelijk	Met veel moeite	Enigszins moeilijk	Weinig moeite	Ja, eenvoudig

B. Oxford Schouder instabiliteit score (Nederlandse versie, versie 2015)[9,11]

Datum: _ _/ _ _ / 20_ _

De test bestaat uit 12 vragen die gaan over (in-)stabiliteit, pijn en mogelijke beperkingen in het persoonlijke en professionele leven van de patiënt.

Een hoge score (hoogste score is 60) betekent een instabiele schouder en/of pijn of beperking in het functioneren ten gevolge van schouderklachten en een lage score (laagste score is 12) betekent een stabiele schouder, weinig pijn of beperking in het functioneren ten gevolge van schouderklachten.

Aangedane zijde		Dominante zijde	

1. Hoe vaak is, **gedurende de afgelopen 6 maanden**, uw schouder uit de kom geschoten?

Niet	1 tot 2 keer in de laatste 6 maanden	1 tot 2 keer per maand	1 tot 2 keer per week	Meer dan 2 keer per week

2. Heeft u, **gedurende de afgelopen 3 maanden**, moeite gehad met of zich zorgen gemaakt om het aantrekken van een T-shirt vanwege uw schouder?

Geen moeite of zorgen	Een beetje moeite of zorgen	Matige moeite of zorgen	Extreme moeite of zorgen	Onmogelijk om te doen

3. Hoe zou u, **gedurende de afgelopen 3 maanden**, de pijn die u aan uw schouder had, zoals die op zijn ergst was, omschrijven?

Geen	Mild	Matig	Hevig	Ondraaglijk

4. In hoeverre werd u, **de afgelopen 3 maanden**, door uw schouder probleem beperkt in het uitvoeren van uw werk (inclusief school, studie, werk of huishoudelijk werk)?

Helemaal niet	Een klein beetje	Matig	Aanzienlijk	Totaal

5. Heeft u, **gedurende de afgelopen 3 maanden**, activiteiten vermeden door zorgen om uw schouder omdat u bang was dat deze uit de kom zou schieten?

Nee, helemaal niet	Heel af en toe	Sommige dagen	De meeste dagen of bij meer dan één activiteit	Elke dag of bij veel activiteiten

6. Heeft u, **gedurende de afgelopen 3 maanden**, vanwege uw schouder probleem activiteiten die belangrijk voor u zijn, niet uitgevoerd?

Nee, helemaal niet	Heel af en toe	Sommige dagen	De meeste dagen of bij meer dan één activiteit	Elke dag of bij veel activiteiten

7. In hoeverre heeft, **gedurende de afgelopen 3 maanden**, uw schouder probleem u belemmerd in uw sociale leven?

Helemaal niet	Af en toe	Sommige dagen	De meeste dagen	Elke dag

8. In hoeverre heeft, **gedurende de afgelopen 4 weken**, uw schouder probleem u belemmerd bij het uitvoeren van sport of hobby's?

Helemaal niet	Een klein beetje / af en toe	Soms	De meeste dagen	Altijd

9. Hoe vaak heeft u, **gedurende de laatste 4 weken**, aan uw schouder gedacht?

Nooit, of alleen wanneer iemand ernaar vroeg	Af en toe	Sommige dagen	De meeste dagen	Elke dag

10. In hoeverre heeft, **gedurende de afgelopen 4 weken**, uw schouder probleem u belemmerd om zware objecten op te tillen of u bereidheid tot tillen beïnvloed?

Helemaal niet	Af en toe	Sommige dagen	De meeste dagen	Elke dag

11. Hoe zou u, **gedurende de afgelopen 4 weken**, de pijn die u gewoonlijk aan uw schouder heeft ervaren, omschrijven?

Geen	Erg mild	Mild	Matig	Hevig

12. Heeft u, **gedurende de afgelopen 4 weken**, vanwege uw schouder, vermeden om 's nachts in bed in bepaalde posities te liggen?

Geen enkele nacht	1 of 2 nachten	Sommige nachten	De meeste nachten	Elke nacht