



RESEARCH ARTICLE

VALIDITY AND RELIABILITY OF THE ARABIC VERSION OF OSS IN ASSESSMENT OF SHOULDER JOINT PAIN

¹, *Mahmoud Mohammed Abdel Rahman, ²Awatef M. Labeeb, and ²Soheir S. Rezk Allah

¹International Medical Center, Ministry Of Defense, Cairo, Egypt

²Basic Science Department, Faculty of Physical Therapy, Cairo University, Giza, Egypt

ARTICLE INFO

Article History:

Received 22nd April, 2018

Received in revised form

20th May, 2018

Accepted 10th June, 2018

Published online 30th July, 2018

Keywords:

Validity,
Reliability,
OSS,
Shoulder joint pain.

ABSTRACT

Background: Oxford Shoulder Score is a shoulder-specific instrument designed to assess the outcome of all shoulder surgeries. It's now used in a number of countries and has also been applied in cohort studies, audits, and national joint replacement registries.

Objective: To test the face validity, the content validity, the feasibility; the internal consistency reliability and the test retest reliability of Arabic-language version of OSS in assessment of shoulder joint pain.

Methods: Forty-three patients with shoulder joint pain caused by degenerative or inflammatory disorders of the Shoulder joint was recruited and 86 sheets (test and retest sheets) were filled out and three expert panels (each consists of ten experts) participated in this study. Forward translation, development of preliminary initially translated version, backward translation, and development of the pre-final version and testing of pre-final version using experts then testing of the final version on patients was done. Index of clarity, expert proportion of clearance, index of content validity, expert proportion of relevance, descriptive statistics, missed item index, Cronbach's alpha and Spearman's rank correlation coefficient were used for statistical analysis.

Results: The study showed that scale index of clarity equals 88.04%, scale index of content validity equals 95.83%, scale-level content validity index universal agreement equals 95.83 %, The scale items were filled by 100% in all sheets. The scale needed less than 5 minutes to be answered in about 98.5%.Cronbach's alpha equals 0.932 (0.897, 0.958) and all Spearman's correlations between test and retest results were statistically significant.

Conclusion: Arabic-language version of the OSS has face and content validity, feasibility and internal consistency and test retest reliability enough in assessment of shoulder joint pain.

INTRODUCTION

Shoulder pain is common chief complain presenting in current orthopedic practices. Shoulder pain is reported to be the third most frequent musculoskeletal symptom after low back and neck pain. In addition, shoulder pain is common in patients with degenerative or inflammatory disorders (Luime *et al*, 2004). Current literature shows that in the operative management of shoulder disorders, including arthroscopic shoulder procedures for rotator cuff, instability, adhesive capsulitis, biceps tendon disease, sub acromion impingement and glenohumeral joint replacement, patient-reported shoulder specific clinical measures are being used widely. This indicates the importance of patients' opinions regarding their health status and quality of life related to their presenting disease in order to assess treatment efficacy (Greving *et al*, 2012). There are many English shoulder measures of shoulder pain were developed such as: Shoulder Pain and Disability Index, The Constant Score, DASH Score, Rating Sheet for Bankart Repair, UCLA Shoulder Score, American Shoulder and Elbow

Surgeons Evaluation Form (ASES), Shoulder Rating Questionnaire, The Western Ontario Shoulder Index and Rotator cuff Quality of Life (RC-QOL) and Oxford shoulder score (OSS) (Kirkley *et al.*, 2003). Oxford Shoulder Score (OSS) is a patient reported questionnaire including 12 descriptors of pain and disability for shoulder ailments. Item rating ranges from 1 to 5 and the total score is from the summation of all 12 rated items from 12 (the best) to 60 (the worst) (Dawson *et al.*, 1996). The OSS was done to translate into Arabic version for Arabic people for assessment of shoulder joint pain. This study considered with the measurement of validity and reliability of the Arabic version of the OSS.

MATERIALS AND METHODS

Participants and design: This study was conducted in outpatient clinics of Al- Qasr Al Aini Hospital and outpatient clinic of El Tahrir hospital and to investigate the validity and reliability of Arabic version of the OSS in patients with shoulder joint pain. This study followed studies that recommended guidelines for translating, adapting and

validating psychological instruments (Borsa *et al.*, 2012; Sousa *et al.*, 2011). Three expert panels; each consists of ten experts and 43 patients with shoulder pain caused by degenerative or inflammatory disorders of the Shoulder joint participated in this study. Each participant signed the consent form.

A. Inclusion criteria for experts

- Experience not less than ten years or at least master degree.
- Major part of their work is with Arabic population.
- Fluent in Arabic and English.

B. Inclusion criteria for patients

All patients were chosen according to the following criteria:

- Referred as shoulder pain such as degenerative or inflammatory disorders of the Shoulder joint by physician.
- Being conscious and ambulant.
- Able to read and write in Arabic.

C. Exclusive criteria for patients

- Pregnant woman.
- Patients with deformity of the upper limb.
- Patients under analgesic medications.
- History of epilepsy.
- Patients with polyneuropathy.
- Bone disease or infection.

Procedures

The following steps were followed:

1- Forward translation: translation of the original scale into Arabic (forward translation or one-way translation) Scale in English was translated to Arabic to produce two forward-translated versions of the scale (A1 and A2).

- Two translators participated in forward translation, their mother language is Arabic, but they had distinct backgrounds
 - One translator was knowledgeable about health terminology and the content area of the construct of the tool in the Arabic.
 - The other translator was knowledgeable about the cultural and linguistic nuances of the Arabic.

2- Development of the preliminary initial translated Arabic version.

Both versions (A1 and A2) were compared and merged by the researchers and research committee of basic science for physical therapy, Faculty of Physical Therapy was asked for help in resolving ambiguities and discrepancies.

3- Blind back-translation (blind backward translation or blind double translation) of the preliminary initial translated version of the scale:

- The preliminary initial translated version of the scale was translated to English to produce two back-

translated versions (B1 and B2).

b) Two translators participated in back translation, but they had distinct backgrounds

- One translator was knowledgeable about health terminology and the content area of the construct of the tool in the English.
- The other translator was knowledgeable about the cultural and linguistic nuances of the English.

4- Comparison of the two back-translated versions of the scale (B1 and B2).The researchers compared back-translation of the scale B1 with B2, and also compared both B1 and B2 with the original English scale

5- Regarding instructions, items, response format, wording, sentence structure, meaning and relevance, and they found that there were no significant differences between them, so the preliminary initial translated Arabic version was considered to be the pre-final Arabic version of the scale.

6- Pilot testing of the pre-final Arabic version of the scale for face and content validity.

a) The first expert panel (ten experts) were asked to evaluate each item of the tool for clarity (face validity) and provide suggestions to improve its clarity; dichotomous questions (clear/unclear) is used regarding instruction (1), items (12) and response words (5) with a total of 29 answer needed from each expert.

b) Then the second expert panel reassessed the clarity of modified pre-final Arabic version of the scale.

c) Then the third expert panel (ten experts) were asked to evaluate each item of the modified pre-final Arabic version of the scale for content equivalence (content-related validity) using the following scale: 1 = not relevant; 2 = unable to assess relevance; 3 = relevant but needs minor alteration; 4 = very relevant and succinct and give suggestions to improve its relevance (1 and 2 considered not relevant, 3 and 4 considered relevant).

d) After the modified pre-final version passed expert face and content validity tests, it was named the final version.

7- Pilot test of the final Arabic version of the scale was conducted on patients with degenerative or inflammatory disorders of the shoulder joint: Patients filled out 86 data collection sheets which was used to collect demographic data (name, age, gender, address, smoking), and OSS.

8- Feasibility (ability to use on larger sample) was evaluated by the assessment of the frequency of missing answers per item and administration time.

9- Eighty six sheets were refilled out again after two days.

Statistical analysis

SPSS computer program (version 20) was used for data analysis:

- Face validity was tested by clarity index and expert proportion of clearance.
- Content validity was tested by index of content validity (CVI) and expert proportion of relevance.

- Descriptive statistics of patients and sheets were made using mean, median, standard deviation (SD), mode, minimum (min) and maximum (max).
- Feasibility index was calculated using missed item index and time taken to fill the questionnaire.
- Internal consistency reliability was measured using Cronbach's coefficient alpha.
- Test retest reliability was measured using mean scores and Spearman's rank Correlation.

RESULTS

Item index of clarity was calculated and it was found that scale index of clarity equals 88.04% and scale level clarity index universal agreement (UA) equals 100% as shown in Table1. Expert proportion of clearance was calculated and represented in Table 2. Item index of content validity of the final version was calculated and it was found that scale index of content validity (S-CVI) equals 100% as shown in Table 3. Also Expert proportion of relevance was calculated and represented in Table 4. Patients were of both genders (30 female and 13 male) also 43 patients made retest. Internal consistency calculations were made for the final version and it was found that Cronbach's alpha equals 0.932 with lower bound 0.897 and upper bound 0.958 at 95% confidence interval.

Table 1. Item index of clarity of the final version

Item	Number of rater's agreements (clear responses)	Item index of clarity
Instructions	10	100%
(1)	9	90%
(2)	8	80%
(3)	9	90%
(4)	8	80%
(5)	9	90%
(6)	8	80%
(7)	8	80%
(8)	7	70%
(9)	8	80%
(10)	9	90%
(11)	8	80%
(12)	9	90%
(13)	10	100%
(14)	9	90%
(15)	9	90%
(16)	10	100%
(17)	9	90%
(18)	9	90%
(19)	10	100%
(20)	10	100%
(21)	10	100%
(22)	8	100%
(23)	10	100%
(24)	7	70%
(25)	9	90%
(26)	7	70%
(27)	9	90%
(28)	8	80%
(29)	8	80%
(30)	9	90%
(31)	8	80%
(32)	9	90%
(33)	8	80%
(34)	8	80%
(35)	9	90%
(36)	9	90%
(37)	10	100%
(38)	10	100%
(39)	10	100%
(40)	10	100%
(41)	9	90%
Mean	8.8	88.04%

(1)→ (12) items of the scale in order, (13) → (41) response words.

Table 2. Expert proportion of clearance of the final version

Expert number	Number of expert agreements (clear responses)	Proportion of clearance
1	39	95%
2	36	87%
3	39	95%
4	36	87%
5	29	70%
6	41	100%
7	40	97%
8	41	100%
9	38	48%
10	40	97%
Mean	8.6	86.07%

Table 3. Item index of content validity of the final version

Item	Number of raters that agree (relevant responses)	I-CVI
(1)	10	100%
(2)	10	100%
(3)	10	100%
(4)	8	80%
(5)	10	100%
(6)	10	100%
(7)	9	90%
(8)	9	90%
(9)	10	100%
(10)	10	100%
(11)	10	100%
(12)	9	90%
Mean	9.5	95.83%

(1)→(12) items of the scale in order, I-CVI: item index of content validity index

Table 4. Expert proportion of relevance of the final version

Expert number	Number of agreements (relevant responses)	Proportion of relevance
1	12	100%
2	10	83%
3	12	100%
4	12	100%
5	12	100%
6	12	100%
7	12	100%
8	12	100%
9	12	100%
10	9	75%
Mean	11.5	95.83%

Table 5. Spearman's rank correlations coefficients

Item No.	r value	Correlation strength	Results of test regarding association between test and retest results
1	0.777	Strong	statistically significant
2	0.725	Strong	statistically significant
3	0.745	Strong	statistically significant
4	0.864	Very strong	statistically significant
5	0.795	Strong	statistically significant
6	0.820	Very strong	statistically significant
7	0.790	Strong	statistically significant
8	0.814	Very strong	statistically significant
9	0.720	Strong	statistically significant
10	0.804	Very strong	statistically significant
11	0.844	Very strong	statistically significant
12	0.819	Very strong	statistically significant
Total score	0.793083	Strong	statistically significant

r: Spearman's rank correlation

Correlations between test and retest results were done. Regarding that the two-tailed value of P is 0.01, Spearman's rank correlations were calculated as shown in Table (5).

DISCUSSION

The present study was designed to test the face validity; the content validity, the Feasibility, the internal consistency reliability and the test retest reliability of Arabic-language version of OSS in assessment of shoulder joint pain. The final Arabic version of OSS has excellent face validity as scale index of clarity equaled 88.04%, and the mean of proportion of clearance (clear responses) equaled 86.07%, also it has excellent content validity as S-CVI equaled 95.83%, S-CVI/UA equaled 95.83% and the mean of the proportion of relevance (relevant responses) equaled 95.83%. The results of the current study came in agreement with Polit Beck, 2006. who stated that a scale to be judged as having excellent content validity, it would be composed of items with item index of content validity (I-CVI) that meet the following criteria (I-CVI of 1.00 with three to five experts and a minimum I-CVI of .78 for 6 to 10 experts) and it would have S-CVI of .90 or higher. The recommended standards may necessitate two rounds of expert review if the initial assessment suggests the need for substantial item improvements (Polit and Beck, 2006). Also this came in agreement with Waltz *et al.* 2005 who stated that S-CVI/Ave of 0.90 or above is the minimum acceptable index, and items that do not achieve the minimum acceptable indices are revised and re-evaluated (Waltz *et al.*, 2005). The Arabic version of OSS has high feasibility because the scale items were filled out by 100% in all sheets and it needed less than 5 minutes to be answered in about 98.5% of all sheets.

The results of the current study came in agreement with Van *et al.* 2015 who stated that Missing rate on the item level was considered acceptable if no single item had a missing rate exceeding 10% and completion time was considered acceptable if 95% of sheets were completed in less than 15 minutes (Van *et al.*, 2015). The Arabic version of OSS has excellent internal consistency and good test retest. Reliability as Cronbach's alpha equaled 0.932 (0.897, 0.958) and all Spearman's rank correlation coefficients between test and retest results were statistically significant (item 1: 0.777, item 2: 0.725, item 3: 0.745, item 4: 0.864, item 5: 0.795, item 6: 0.820, item 7: 0.790, item 8: 0.814, item 9: 0.720, item 10: 0.804, item 11: 0.844, item 12: 0.819, total score: 0.79308). So according to George and Mallery 2003 α above 0.9 is referred as excellent internal consistency, also Spearman's rank correlation coefficient between 0.7 and 0.9 (as in all items) is referred as good test retest reliability (George and Mallery, 2003). These results came in agreement with a study that conducted to assess its validity in native French-speaker patients with shoulder pain. The translation process was carried out following a translation/back-translation methodology by two translators. One hundred forty-four patients suffering from degenerative or inflammatory diseases of the shoulder were included in the study. The average time required to complete the French OSS was 2 min and 45 s. seventy patients were asked to complete the questionnaire twice (test/retest reliability). Internal consistency was high with Cronbach's α coefficient = 0.93. The interclass correlation coefficient was 0.91 (95% CI: 0.88–0.94) for test/retest reliability. The French OSS score was significantly correlated with the Constant–Murley score ($r = 0.73$ and $P < 0.0001$) and with the SSV ($r = 0.68$ and $P < 0.0001$). The study shows that the French version of the OSS is reliable, valid, and reproducible. The score was adapted to the French-speaking population for the self-assessment of patients with degenerative or inflammatory disorders of the Shoulder. (Tuton *et al.*, 2016). Also these results came in agreement with

similar results obtained by Luigi Murena *et al.* 2010 who conducted a study to adapt the OSS to the Italian version and to translate, and validate the Italian version of the OSS. They recruited 140 patients with shoulder pain caused by degenerative or inflammatory state or disorder of the shoulder. Patients completed the following questionnaires: Italian OSS, University of California, Los Angeles (UCLA) Shoulder Rating Scale, Constant-Murley shoulder assessment, and the Medical Outcome Study Short-Form 36 Health Survey (MOS SF-36).

Internal consistency was tested using Cronbach coefficient α . Reproducibility was assessed by asking 110 patients to complete another OSS 48 hours after the first. Correlation between the total results of both tests was determined by the Pearson correlation coefficient. Validity was assessed by calculating the Pearson correlation coefficient between the OSS and the UCLA, Constant-Murley, and SF-36 assessments. Reproducibility was assessed by asking 110 patients to complete another OSS 48 hours after the first. Correlation between the total results of both tests was determined by the Pearson correlation coefficient. Validity was assessed by calculating the Pearson correlation coefficient between the OSS and the UCLA, Constant-Murley, and SF-36 assessments. Cronbach alpha was 0.95. The Pearson correlation coefficient was $r = 0.97$. With respect to validity, there was a significant correlation between the Italian OSS and the individual scores of UCLA, Constant-Murley, and SF-36. Psychometric properties of the Italian OSS compared well with those reported for the English OSS. As demonstrated by the high values of Cronbach α and Pearson correlation coefficients, in accordance with the English version of the OSS, the Italian version proved to be a reliable, valid, and reproducible measure of shoulder pain perception in Italian-speaking patients (Luigi Murena *et al.*, 2010). Other study aimed to translate and culturally adapt a Turkish version of the OSS and validate its use for assessing Turkish patients with shoulder pathology. Eighty-four patients (mean age 49.26 ± 11.92 years) with Shoulder problems participated. Patients completed the Turkish OSS, the Short Form 36 (SF-36), and the Shoulder Pain and Disability Index (SPADI). Internal consistency was tested using Cronbach α coefficient. Reproducibility was assessed by asking patients to complete another OSS 48 h after the first test. Correlation between the total results of both tests was determined by the Pearson correlation coefficient and ICC. Validity was assessed by calculating the Pearson correlation coefficient between the OSS and SPADI and SF-36 scores. The internal consistency was high (Cronbach's $\alpha = 0.92$). The reproducibility tested by two different methods showed no significant difference. Correlation between the OSS and SPADI and SF-36 physical component summary score were -0.7, and 0.6, respectively ($p < 0.001$). There was no floor or ceiling effect in total OSS score. The Turkish version of the OSS proved to be valid, reliable and reproducible instrument as demonstrated by high Cronbach α and Pearson. Correlation Coefficients. The application and evaluation of the instrument was feasible and minimally time consuming for use in clinical trials in Turkish-speaking patients with shoulder problems. (Tuğay *et al.*, 2011). Validity and reliability of translated tools were made over two or three studies not one. The first study is designed to translate the tool to the targeted language then test the translated version for face and content validity then test the reliability; it was conducted on monolingual population. The second study was designed to test the full the psychometrics of the translated tool with bilingual participants. The third study

is conducted to test the full psychometric properties of the translated tool on monolingual population, noting that the second study is not necessary to be made (Borsa *et al.*, 2012; Sousa *et al.*, 2011). This study is considered to be the first study in the validity and reliability studies of the Arabic language version of OSS. The final version is considered the base for the next research that will be conducted to establish the full psychometric properties of Arabic- language version of OSS.

Conclusion

The results obtained from the current study and the discussion that followed it can lead to concluding that Arabic-language version of the OSS has face and content validity, feasibility and internal consistency and test retest reliability enough in assessment of shoulder joint pain.

Acknowledgments: The researchers are grateful to the volunteers and participants involved in this study.

REFERENCES

- Luime JJ, Koes BW, Hendriksen IJ, Burdorf A, Verhagen AP, Miedema HS, et al. 2004. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol.*, 33:73-81.
- Greving K, Dorrestijn O, Winters JC, Groenhof F, van der Meer K, Stevens M, et al. 2012. Incidence, prevalence, and consultation rates of shoulder complaints in general practice. *Scand J Rheumatol.*, 41:150- 5.
- Kirkley A, Griffin S, Dainty K. 2003. Scoring systems for the functional assessment of the shoulder. *Arthroscopy: Journal of Arthroscopic and Related Surgery*, 19(10): 1109-1120.
- Dawson J, Fitzpatrick R, Carr A. 1996. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br.*, 78:593-600.
- Borsa J, Damásio B, Bandeira D. 2012. Cross-cultural adaptation and validation of psychological instruments: Some considerations. *Paidéia (Ribeirão Preto)*, 22:423-32.
- Sousa V, Rojjanasrirat W. 2011. Translation, adaptation and validation of instruments or scales for use in cross cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract.*, 17:268-74.
- Polit D. and Beck C. 2006. The content validity index: are you sure you know what's being reported? Critique and recommendations. *Nurs Health*, 29(5):489-497.
- Waltz C, Strickland O, Lenz E. *Measurement in Nursing and Health Research*, 3rd edition. New York: Springer Publishing Company; 2005.
- Van V, Birnie E, Poeran J, Torij H, Steegers E, Bonsel G. Feasibility and reliability of a newly developed antenatal risk score card in routine care. *Midwifery* 2015; 31:147-54.
- George D. and Mallery P.: *SPSS for Windows step by step: A simple guide and reference*, 11.0 update, 4th edition. Boston: Allyn and Bacon, 2003.
- Tuton D., C. Barbe, J.-H. Salmon, M. Dramé, C. Nérot, X. Ohl, 2016. Transcultural validation of the Oxford Shoulder Score for the French-speaking population.
- Luigi Murena, MD, Ettore Vulcano, MD, Fabio D'Angelo, MD, Maria Monti, MS, Paolo Cherubino, MD. 2010. Italian cross-cultural adaptation and validation of the Oxford Shoulder Score.
- Tuğay U, Tuğay N, Gelecek N, Özkan M. 2011. Oxford Shoulder Score: cross-cultural adaptation and validation of the Turkish version.
