

RESEARCH

Open Access



Oxford Shoulder Instability Score: cross-cultural adaptation into Spanish and analysis of its methodological quality

Rocio Aldon-Villegas¹, Gema Chamorro-Moriana^{1*}, Fernando Espuny-Ruiz^{1*} and Maria-Luisa Benitez-Lugo¹

Abstract

Background Oxford Shoulder Instability Score (OSIS) is a patient reported outcome measure designed specifically to assess functional difficulties resulting from shoulder instability. The main aim of this study was to cross-culturally adapt the OSIS to Spanish. Secondary, it aimed to analyse its methodological quality.

Methods A cross-cultural adaptation to Spanish has been carried out following the recommendations of Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) with a sample of 167 cases of shoulder instability. Inclusion criteria were: subjects with instability symptoms in at least one shoulder with or without clinical diagnosis; aged between 18 and 60. The following psychometric properties were evaluated: validity (construct, internal and external), reliability (internal consistency, test-retest and measurement error), discriminant ability and feasibility. The methodological quality was addressed with Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) and COSMIN Risk of Bias checklist (COSMIN RoB).

Results The construct validity obtained an OSIS correlation with the Simple Shoulder Test of $r=0.636$ and with the Western Ontario Shoulder Instability of $r=0.80$. The unidimensionality of the OSIS was confirmed through second-order factor analysis. Cronbach's alpha and intraclass correlation coefficient were both 0.93 (IC95%: 0.91–0.94). Standard error of measurement was 0.70, and the percentage of error and smallest detectable change were 1.46% and 1.94, respectively. No floor or ceiling effects were found. Assessing feasibility of OSIS, the participants answered all questions, had no questions and completion time was: mean 2 min 30 s; $SD \pm 1$ min. Regarding methodological quality, the study showed *low* risk of bias in the areas of patient selection, index test, reference standard and flow and timing, as well as *low* concern regarding applicability in the domains of patient selection, index test and reference standard according to QUADAS-2. In relation to COSMIN RoB, all psychometric properties, except for content validity, obtained *very good* results.

Conclusions The Spanish version of OSIS offers valid, reliable and feasible functional outcome measures for Spanish-speaking subjects with shoulder instability, based on its psychometric properties and its methodological quality.

*Correspondence:
Gema Chamorro-Moriana
gchamorro@us.es
Fernando Espuny-Ruiz
f.espuny_98@hotmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Keywords Shoulder instability, Cross-cultural adaptation, Oxford shoulder instability score, Psychometrics properties, Methodological quality

Background

Shoulder instability is a dysfunction characterized by abnormal translation of the humeral head relative to the glenoid cavity, impairing shoulder comfort and function [1]. The shoulder is the joint that is most susceptible to instability [2] due to its particular biomechanics. In fact, approximately 45% of all bodily luxations correspond to the glenohumeral joint [3]. Its incidence is between 25 and 47 cases per 100,000 inhabitants per year in developed countries [4], commonly affecting active [5] and young people, between the second and third decade of life [6]. Consequently, this dysfunction is a current problem in orthopedic practice [7].

The diagnosis of shoulder instability can be confusing due to inadequate assessment [8]. For this reason, many functional assessment scales have been developed and used in clinical and research settings [9]. A great number of them have proven to be reliable, valid and effective assessment methods. Its use has increased in recent years [10], as it enables therapists and patients to work with a greater degree of objectivity, unify the language of clinicians, create treatment protocols and compare results. Within these scales, there is a clear trend towards the development and use of Patient Reported Outcome Measures (PROMs) [11] in order to quantify the patient's health condition without interpretation by clinicians [12]. Many PROMs were designed for general shoulder assessment [13], and even some of them for specific dysfunctions [13], such as the Oxford Shoulder Instability Score (OSIS) [14].

OSIS, created by Dawson et al. in 1999 [14], proving to be a valid, reliable tool, sensitive to changes of clinical importance, as well as practical and short, i.e. easy and quick to fill in [14]. It was originally created and validated in English [14]. However, the interest aroused by this scale in clinical and research fields is leading to its cross-cultural adaptation. This process addresses both linguistic issues (translation) and cultural adaptation [15], resulting in versions in other languages, such as Dutch [16], Italian [7], Persian [17], Turkish [18], Arabic [19] and Norwegian [20]. Since Spanish is the second most spoken language in the world, it would be convenient to carry out its cross-cultural adaptation. This would facilitate its use in Spanish-speaking contexts, as well as comparison with other populations, and would enable scientists to learn about its psychometric properties [21]. In addition, it would be desirable that all adaptations adequately address its psychometric properties and offer optimal methodological quality. This is the only way to promote reproducibility, i.e., the ability to obtain consistent results

when a study is repeated under the same conditions [22], and to ensure transparency of the results, avoid biases, etc., providing clinicians with confidence in their assessments and, therefore, in their decision making.

Therefore, the main objective of this study was to cross-culturally adapt the Oxford Shoulder Instability Score to Spanish. A second objective was to analyze the methodological quality of this adaptation.

Methods

Study design

Considering the new COSMIN recommendations regarding methodological quality [23], this study undertook the process of cross-cultural adaptation of OSIS to Spanish after approval by Carr, coauthor of the original OSIS study [14]. In addition, this article analyzed the methodological quality of such adaptation. This study was approved by the Ethical Committee of the Andalusian Public Health System (No. 1267-N-21) in accordance with the Helsinki Declaration.

Translation and cultural adaptation procedure

The forward-backward translation and cultural adaptation process from the original English version to Spanish of OSIS, followed the guide by Beaton et al. [15]:

1. Two bilingual translators whose mother tongue is Spanish produced two independent translations from the English to the Spanish language. One was an expert in translation and interpretation and the other in musculoskeletal disorders. Emphasis was placed on translations being semantic, prioritizing conceptual equivalence. As a result, two versions of the original OSIS scale were obtained in Spanish.
2. Synthesis: two experts in shoulder instability evaluated the two versions and, together with the translators, agreed on a final version.
3. Back translation: it was performed independently by two native English speakers fluent in Spanish, who were blinded from the original version. One of them had clinical knowledge of shoulder problems. Both reached an agreement on the back-translated version.
4. Back translation review: a group of four experts, consisting of the translators involved and the two clinical experts mentioned above, reviewed the agreed back-translation. The experts compared the new version with the original.
5. Pilot testing: A pilot test of the final Spanish version of OSIS with a sample of 20 shoulders with instability

belonging to 20 participants was conducted. The comprehension and/or clarity of the questionnaire, as well as the appropriateness of the vocabulary and the relevance of the expressions in the Spanish culture were noted and assessed.

6. Validation of the Spanish version of OSIS with a sample of 167 shoulders with instability belonging to 107 participants. The following were assessed: validity: construct, internal, and external; reliability: internal consistency, test-retest, and measurement error; discriminant ability; and feasibility (see section on assessment of psychometric properties).

Participants and sample

Participants were recruited using non-random consecutive sampling from University, clinical and sports centres in Seville and Cadiz. Patients were collected through dissemination in these institutions (posters, mails, contacts, etc.) where people with a feeling of instability, i.e. with the perception that their shoulder could come out, with or without previous dislocation/subluxation, whether or not they had undergone surgery, were requested. This perception was confirmed by medical history. Therefore, inclusion criteria were: (1) subjects with instability symptoms in at least one shoulder with or without clinical diagnosis, i.e. whether or not diagnosed by a clinician; (2) aged between 18 and 60.

Excluded: (1) participants with musculoskeletal pathologies not associated with possible instability (e.g., humeral neck fracture, tendon rupture, rheumatoid arthritis, brachial paralysis); (2) people with cognitive impairment that rendered them unable to complete the functional assessment tools.

The study considered the affected shoulders as the sample, except for the descriptive analysis, where the sample was the participants.

Patient reported outcome measures used

Questionnaire to be adapted to Spanish: OSIS

OSIS is a disease-specific PROM that was developed for assessing shoulder instability (1999) [14]. It is composed of a total of 12 items that address the following topics: instability (1), work (1), pain (2), activities of daily living (4), physical and sport activities (1), social life (1) and psychological aspects (2). There are two scoring systems: the original system, in which each item ranged from 1 to 5 points, with the total score ranging from 12 (smallest disorder) to 60 (highest degree of disability) [14]; and the scoring system updated by the OSIS authors themselves [24], in which each item scored from 0 to 4 with the total score ranged from 0 (most impaired) to 48 (least impaired), i.e., the score was reversed. This study employed this new scoring system.

The cross-cultural adaptations to Dutch [16] and Italian [7], Arabic [19] used the current scoring system, while the Turkish [18], Persian [17] and Norwegian [20] versions used the original.

Spanish adapted questionnaires used for validation: Western Ontario Shoulder Instability (WOSI) [25], Simple Shoulder Test (SST) [26]

WOSI is a disease-specific PROM that was developed by Kirkley et al. (1998) [27] for assessing shoulder instability. It contains 21 items that assess 4 domains: physical symptoms (10 items), sport and recreation work (4 items), lifestyle (4 items), and emotions (3 items). Each item ranged from 0 to 10 and the total score ranged from 0 (optimal shoulder condition) to 210 points (highest degree of disability).

SST was validated in shoulder instability by Godfrey et al. (2007) [28], although it was originally designed by Lippitt et al. (1993) [29] for shoulder issues in general. It is composed of 12 items about activities of daily living (3 items), range of movement (3 items), strength (3 items), physical and sport activities (2 items) and work (1 item). It is a dichotomous scale with possible responses of “yes” or “no” where each positive answer sums one point. Thus, the total score ranged from 0 (highest disability) to 12 points (optimal shoulder condition).

Protocol for action

After reading the study information sheet and the informed consent form, and offering the opportunity to answer any questions, all participants provided their written consent.

After completing the content validation corresponding to the process of translation and cross-cultural adaptation into Spanish, two researchers (RA and MB) familiar with the use of the shoulder-specific PROMs, administered the questionnaires in 2023, after sharing their instructions. One of them (RA) was present at all times to ensure that all participants completed the scales in their entirety and under the same conditions.

The fieldwork was conducted from February to July 2023. It was undertaken in equipped facilities in Cadiz and Seville, where it was confirmed by medical history that the participants had a feeling of instability. Descriptive data were collected –affiliation, date of birth, sex, weight, height, body mass index, affected shoulder(s), dominant arm(s), physical activity, education level, occupation, treatment received for said dysfunction– and the SST, WOSI and OSIS questionnaires were completed, in that order. All subjects had the opportunity to ask questions.

Participants completed the aforementioned questionnaires at two points in time with an interval of 7–14 days. This interval was considered long enough to forget prior

answer, and short enough to assume an unchanged shoulder condition [30]. The completion time was recorded without taking into account the time for clarifications. Participants with bilateral shoulder instability answered for each shoulder independently.

Analysis of psychometric properties

Construct validity (convergent), internal and external validities, internal consistency, test-retest reliability, measurement error, discriminant ability (floor and ceiling effects) and feasibility were assessed following COSMIN guidelines [31].

Construct validity. The authors formulated pre-established hypotheses regarding the expected correlation of Spanish OSIS with SST and with WOSI. These correlations are based on the experience of other researchers [16] and in our own knowledge of these PROMs. The hypotheses were as follows:

1. Correlation between Spanish OSIS and SST of at least $r = 0.6$.
2. Correlation between Spanish OSIS and WOSI of at least $r = 0.7$.

Note that criterion validity could not be assessed because PROMs assess constructs that can only be reported by patients, and therefore, there are no gold standards for these measures [32].

Internal and external validity. Internal validity was analyzed by means of an analysis of the factor structure of the tool. External validity refers to the ability of a questionnaire to measure accurately and under different conditions (subjects, circumstances, times). Thus, this study included a diverse sample in terms of age and professional-socio-cultural characteristics, from urban and rural areas, in different seasons of the year (winter, spring and summer).

Reliability. It included the assessment of: internal consistency; temporal stability of the results by the test-retest method (washout period 7–14 days); and measurement error, i.e. the systematic, random error in the construct, not attributed to true changes in patient's condition [33]. This error can be expressed as the standard deviation of repeated measurements in a single participant, referred to as the standard error of measurement (SEM). The error percentage was used to obtain the SEM in relative terms. The smallest detectable change (SDC) represents the minimal change that a patient must show to ensure that the observed change is real, and not a measurement error [33]. SEM, error percentage and SDC were calculated.

Discriminant ability. Floor and ceiling effects were analyzed by calculating the percentage of shoulders with

the highest and lowest scores (10th and 90th percentiles, respectively) for the overall scores [34].

Feasibility. The following were considered in evaluating the feasibility of OSIS: the presence of unanswered questions, the questions of the participants and the time taken to complete the questionnaire.

Statistical analyses

IBM SPSS STATISTICS 27 software for analyses. $p < 0.05$ values were considered statistically significant.

Regarding the sample size (shoulders), for a confidence level of 95%, i.e., an alpha type error of 5% and a precision of 3.5%, at least 123 shoulders are required [35].

As for the descriptive data of the participants and the sample, the absolute (N) and relative (%) frequencies were considered for the qualitative variables. For quantitative variables, normality was assessed using the Kolmogorov-Smirnov for $n \geq 50$ samples or Shapiro-Wilks test for $n < 50$ [36]. The mean and standard deviation were considered for parametric variables, and the median and interquartile range for nonparametric variables.

Construct validity was assessed using Pearson's or Spearman's correlation coefficient, depending on the distribution of the data. Its interpretation was according to Martinez-Ortega RM et al. [37]: $r \leq 0.25$, *rare*; $0.25 \geq r > 0.5$, *weak*; $0.5 \geq r > 0.7$, *moderate* and $r > 0.7$, *strong* correlation.

Internal validity. The Kaiser-Meyer-Olkin test (KMO) for sampling adequacy and Bartlett's sphericity test (BST) were performed to measure homogeneity of variance. This is done to evaluate whether the data are suitable for factor analysis. KMO value should be close to 1 and BST value should be significant ($p < 0.05$). An exploratory factor analysis was performed using the main component extraction method with varimax rotation. They are considered acceptable when the proportion of variance explained is $> 50\%$ [38]. The factorial rotation matrix was calculated, estimating acceptable values > 0.5 . Items with a < 0.40 load on both factors (crossloading factors) were assigned to the factor with the highest correlation [39].

Reliability. The internal consistency was assessed with the Cronbach's alpha that ranges from 0 to 1, and values ≥ 0.7 indicate adequate internal consistency for scale [40]. The test-retest reliability was tested using intraclass correlation coefficient (ICC). ICC values were interpreted following Fleiss guidelines [41]: < 0.4 , *poor*; $0.4 - 0.75$, *good*; > 0.75 , *excellent*. SEM was calculated: Standard Deviation (SD) $\times \sqrt{1 - \text{ICC}}$. Error percentage was calculated: $(\text{SEM} / \text{average}) \times 100$. SDC defined as $1.96 \times \sqrt{2\text{SEM}}$ [33].

Discriminant ability. Ceiling and floor effects were calculated using the percentage of shoulders with the highest and lowest scores. These effects were considered to be present when more than 15% were in these percentiles [34].

Assessment of methodological quality

Two assessment scales were employed to evaluate the methodological quality of this study: the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) [42] and the Consensus-based Standards for the selection of health Measurement Instruments Risk of Bias checklist (COSMIN RoB) [31].

QUADAS-2 [42] is an evaluation scale for the diagnostic criteria of validation studies. It was employed to evaluate risk of bias and applicability. The seven items of this tool helped identify a *low*, *high* or *unclear* risk of bias in each domain, as well as concerns regarding applicability.

COSMIN RoB [31] was used as a tool to assess the adequacy and validity of the validation studies [43]. This checklist is composed of ten boxes corresponding to ten metric properties, each of which contains items related to design and statistical methods. Each assessment was classified in a range of four levels: *very good*, *adequate*, *doubtful* and *inadequate* [31]. COSMIN RoB [31] does not negatively assess the absence of metric properties that have not been applied in papers.

Results

Participants and sample

The sample consisted of 167 shoulder instability, exceeding the 123 shoulders that are the required sample size for a 95% confidence level, the 120 shoulders that Hair et al. [38] suggest (10 subjects per item) for translations, cultural adaptations and validations; and even exceeding COSMIN recommendations [32] (optimum ≥ 100). 107 participants were involved, of whom 60 were bilaterally affected. Tables 1 and 2 show the descriptive data of the participants and the sample (shoulders).

Validity

Table 3 shows the correlation analysis between OSIS, SST and WOSI to assess construct validity.

The correlation of OSIS with SST and with WOSI was $r = 0.636$ and $r = 0.80$, respectively. Statistically significant correlation among the 3 questionnaires was obtained ($p < 0.05$). In fact, all of them were $p < 0.001$. All their correlation coefficients were > 0.636 .

Tables 4 and 5 show the results of the KMO and BST tests, and the rotated component matrix respectively to assess internal validity.

The percentage of total explained variance is 66.053%.

The unidimensionality of the OSIS was confirmed by performing the second-level factor analysis.

Reliability

Internal consistency and reproducibility: Cronbach alpha was 0.93 (IC95%: 0.91–0.94) and ICC was 0.93 (IC95%: 0.91–0.94) in the same way. Test-retest analysis was statistically significant ($p < 0.001$).

SEM of OSIS was 0.70, percentage error was 1.46% and SDC was 1.94.

Discriminant ability

No ceiling or floor effects were observed. 12% of the shoulders achieved the highest scores and 10.2% obtained the lowest scores.

Feasibility

The mean time, minutes (min) and seconds (s), to complete OSIS was 2 min 30 s (± 1 min SD); SST, 2 min 04 s (± 45 s SD); and WOSI, 5 min 43 s (± 1 min 53 s SD).

All participants completed all OSIS questions without difficulties or queries.

Assessment of methodological quality

The results of the QUADAS-2 [42] and the COSMIN RoB [31] are presented in Table 6.

This study obtained *low* risk of bias in the domains of patient selection, index test, reference standard and flow and timing, as well as *low* concern regarding applicability in the areas of patient selection, index test and reference standard according to QUADAS-2.

In relation to the results of COSMIN RoB, content validity was *doubtful*. Internal consistency, reliability, measurement error and construct validity were *very good*.

Discussion

This study addressed the cross-cultural adaptation of the Spanish version of the OSIS, complemented with an analysis of its methodological quality. The most relevant findings were *strong* validity, *very good* internal consistency, *excellent* test-retest reliability, *good* discrimination capacity, and easy and quick completion, which makes the Spanish OSIS suitable for the assessment of shoulder instability. Regarding methodological quality, this adaptation showed a *low* risk of bias with QUADAS-2 and in general *very good* results in the metric properties evaluated with COSMIN RoB.

The two translations and back-translations during the linguistic adaptation process were undertaken without difficulty. The final Spanish version was free of cross-cultural inconsistencies, so the questionnaire was applicable to the Spanish-speaking population.

Validity, reliability, discriminant ability and feasibility are then discussed.

Validity

Regarding construct validity, a *strong* correlation was found between OSIS and WOSI ($r = 0.80$). Similar results were reported in the studies by Skare et al. [20] and Van der Linde et al. [16], which also indicated *strong* correlations of $r = 0.73$ and $r = 0.82$, respectively, between these two PROMs.

Table 1 Descriptive characteristics of the participants

Participants			Men (n = 72)	Women (n = 35)	Total (n = 107)
Age (years)		$\bar{x}$	29.03	31.66	29.9
		SD	9.16	12.04	10.21
		Med	25.5	28	26
		IQR	23–32.75	22–43	22–34
Weight (kg)		$\bar{x}$	81.43	62.69	75.30
		SD	9.93	9.42	13.14
		Med	80.25	61	75
		IQR	74–85	57–67	65–83.5
Height (cm)		$\bar{x}$	178.69	164.14	173.93
		SD	5.59	5.36	8.79
		Med	179.50	165	175
		IQR	175–182	159–168	168–181
BMI		$\bar{x}$	25.48	23.29	24.76
		SD	2.61	3.50	3.09
		Med	25.1	22.58	24.62
		IQR	23.52–26.56	21.30–24.45	22.66–25.96
Impairment	Unilateral	N	32	15	47
		%	43.66	44.44	43.93
	Bilateral	N	40	20	60
		%	56.34	55.56	56.07
Impaired limb*	Dominant	N	13	8	21
		%	40.6	53.3	44.7
	Non- dominant	N	19	7	26
		%	59.4	46.7	55.3
Physical activity (overhead movements)**		N	66	26	92
		%	91.7	74.3	86
Educational level	Primary	N	0	1	1
		%	0	2.9	0.9
	Secondary	N	19	13	32
		%	26.4	37.1	29.9
	University	N	53	21	74
		%	73.6	60	69.2
Occupation	Management/Professional	N	16	10	26
		%	22.2	28.6	24.3
	Service	N	19	8	27
		%	26.4	22.9	25.2
	Office work	N	4	3	7
		%	5.6	8.6	6.5
	Student	N	33	14	47
		%	45.8	40	43.9
Area of origin	Urban	N	60	31	91
		%	83.3	88.6	85
	Rural	N	12	4	16
		%	16.7	11.4	15

Abbreviations: $\bar{x}$, mean; SD, standard deviation; Med, median; IQR, interquartile range; BMI, body mass index

*Only unilateral affected subjects considered. **Physical activity involving shoulder movements that place the hand above the head

In contrast, a *moderate* correlation was found in this study between OSIS with SST ($r=0.63$). This is consistent with the findings of Van der Linde et al. [16], who reported a *moderate* correlation with SST ($r=0.69$). The higher correlation of OSIS with WOSI could be justified, since both are two specific shoulder instability

questionnaires, whereas SST was designed for shoulders in general [29], although it was validated on unstable shoulders [28].

Several studies, such as those by Dawson et al. [14], Mazzoni et al. [7] and Skare et al. [20], have used, in addition to PROMs, other tools including a clinician's

Table 2 Descriptive characteristics of the sample

Shoulders (sample)		Men (n = 112)	Women (n = 55)	Total (n = 167)
Surgical treatment	N	17	6	23
	%	15.2	10.9	13.8
Conservative treatment	N	63	25	88
	%	56.3	45.5	52.7
OSIS*	$\bar{x}$	34.07	34.02	34.05
	SD	9.43	10.94	9.92
	Med	35.5	36	36
	IQR	29–42	24–44	28–43
SST*	$\bar{x}$	10.40	9.09	10
	SD	1.86	2.77	2.28
	Med	11	10	11
	IQR	10–12	6–11	9–12
WOSI*	$\bar{x}$	80.33	87.5	82.78
	SD	47.11	63.82	53.32
	Med	82.5	81	82
	IQR	38.25–108	22.5–146.3	36–120.5

Abbreviations: $\bar{x}$, mean; SD, standard deviation; Med, median; IQR, interquartile range; OSIS, Oxford Shoulder Instability Score; SST, Simple Shoulder Test; WOSI, Western Ontario Shoulder Instability

*Note: the retest moment was considered for the questionnaires statistics

Table 3 Correlation among OSIS, SST and WOSI

r	SST	OSIS
OSIS	0.636 ($p < 0.001$)	
WOSI	0.638 ($p < 0.001$)	0.800 ($p < 0.001$)

Abbreviations: r, correlation coefficient; OSIS, Oxford Shoulder Instability Score; SST, Simple Shoulder Test; WOSI, Western Ontario Shoulder Instability

Note: all coefficients were analyzed using Spearman's Rho

Table 4 Kaiser-Meyer-Olkin test and Bartlett's sphericity test

Kaiser-Meyer-Olkin measure of sampling adequacy			0.933
Bartlett's sphericity test	Approx. Chi-square		1379.854
	gl		66
	Sig.		0.000

Abbreviations: approx., approximate; gl, degrees of freedom; sig. significance

Table 5 Rotated component matrix^a

	Component	
	1	2
Item 9	0.855	
Item 8	0.840	
Item 5	0.743	
Item 10	0.710	
Item 6	0.696	
Item 1	0.347	
Item 4		0.829
Item 7		0.814
Item 2		0.783
Item 3		0.655
Item 11		0.627
Item 12		0.523

Extraction method: main component analysis. Rotation method: Varimax with Kaiser normalization. a. Rotation has converged in 3 iterations

assessment to assess the construct validity of OSIS. This is the case of the Rowe scale [44]. In contrast to Skare et al. [20], who reported a *weak* correlation between OSIS and Rowe, Dawson et al. [14], Mazzoni et al. [7] found a *moderate* correlation. Notably, Dawson et al. [14] observed a *strong* correlation at a second point in time. These high correlations are particularly relevant, since they make it possible to dispense with the clinician's assessment, which would considerably reduce the time and human resources used in the assessment. This favors the trend towards the development and use of PROMs. Moreover, by focusing on the patient's perception, the effectiveness of personalized treatments is improved.

Regarding findings associating shoulder instability with quality of life [7, 14, 16, 18], Van der Linde et al. [16] correlated OSIS with the subscales of the SF-36 health questionnaire, obtaining a *strong* correlation with the "pain" subscale ($r = 0.78$) and a *moderate* correlation ($r = 0.69$) with the "role-physical" subscale. These results were similar to those described by Dawson et al. [14] in their original validation. This confirms the clear relationship between quality of life and shoulder instability, the symptoms [45] of which imply a detriment in this quality of life.

The internal validity of the Spanish OSIS proved to be high, supported by a KMO index of 0.93, indicating a high level of sampling adequacy. This suggested that the correlations between items were small enough for factor analysis to be appropriate. As for the dimensions of the questionnaire, Van der Linde et al. [16] concluded that OSIS was not unidimensional, because the one-factor model did not fit the data well. In contrast, our second-level factor analysis confirmed the unidimensionality of the OSIS. Similarly, Sonmezer et al. [18] supported the one-factor structure of this scale. This ensured that the questionnaire measured a single construct and that it reflected a correct assessment of shoulder instability [46].

On the other hand, the inclusion of a heterogeneous sample in this study is crucial to ensure external validity. By including participants of different ages, anthropometric, professional, social and geographic characteristics, as well as considering different seasons of the year, we increase the representativeness of the study and reduce possible biases. Consequently, the value of OSIS in different contexts and populations was evident for both clinical and research settings.

Reliability

The results obtained showed *high* internal consistency ($\alpha = 0.928$) and *excellent* test-retest reliability ($ICC = 0.928$). As for internal consistency, previous validation studies [14, 18–20] obtained values between 0.86 and 0.95, being considered *adequate* when ≥ 0.7 . This study showed a high value (0.93), indicating strong

Table 6 Assessment of the methodological quality with QUADAS-2 and COSMIN RoB

QUALITY ASSESSMENT OF DIAGNOSTIC ACCURACY STUDIES-2	RISK OF BIAS				APPLICABILITY		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
							
CONSENSUS-BASED STANDARDS FOR THE SELECTION OF HEALTH MEASUREMENT INSTRUMENTS RISK OF BIAS CHECKLIST	Measurement properties*						
	Content validity	Internal consistency	Reliability	Measurement error	Construct validity		
	Doubtful	Very good	Very good	Very good	Very good		

*Note: Only the metric properties evaluated in this study have been included.

consistency among the scale items. Therefore, it could be generalized that the reliability of the OSIS is adequate.

Van der Linde et al. [16], Hanna et al. [19], Skare et al. [20] and Olyaei et al. [17] confirmed our *excellent* test-retest reliability with an ICC of 0.87, 0.89, 0.90 and 0.94, respectively. Olyaei et al. [17] obtained the best value of the three. However, the short washout period (2 days) and the small sample size ($n=45$) used are a handicap.

Discriminant ability

Unlike the OSIS Italian adaptation study [7], which showed a ceiling effect, the Dutch [16], Arabic [19], Norwegian [20] and this Spanish version showed neither a ceiling nor a floor effect. The absence of these effects supports the content validity of the OSIS [20] and its ability to capture variability in participants' responses, i.e. to distinguish between different levels of impact.

Feasibility

The full clarity of the questionnaire avoided the participants from having questions, who completed it with ease.

The mean time to complete OSIS was 2 min 30 s (± 1 min SD), similar to that used for SST 2 min 04 s (± 45 s SD) and significantly lower for WOSI 5 min 43 s (± 1 min 53 s SD). The fast completion of OSIS makes it an agile tool for data collection.

Assessment of methodological quality

QUADAS-2 [42] and COSMIN RoB [31] were complemented to assess and establish the methodological quality of this cross-cultural Spanish adaptation of OSIS.

As for QUADAS-2 [42], it obtained the best possible result in all its domains which consequently led to a significant decrease in the probability of bias. On the one hand, this adaptation relied on a consecutive sample of unstable shoulders, avoiding convenience sampling. On the other, the results of OSIS, "Index test", and those of WOSI, "Reference standard", were interpreted independently, increasing their accuracy. Furthermore, regarding

the "Flow and timing" domain, both questionnaires were applied at the same time, ensuring that they were completed by all participants, who were included in the statistical analysis.

Regarding the COSMIN RoB [31], the only psychometric property that did not reach an *adequate* level was content validity. Despite having an adequate methodology to assess the relevance, comprehensiveness and understanding of OSIS from the participant and practitioner perspectives, and the inclusion of experts from several disciplines, among others, the limitation was that it was only undertaken on 20 participants, when at least 30 are needed. Internal consistency, reliability, measurement error, and construct validity achieved *very good* results, thanks to the comprehensive methodological design and the appropriate selection of statistics to evaluate each property. This confirms the high methodological quality of the Spanish version of OSIS.

Regarding the limitations of the study, sensitivity was not calculated due to the absence of intervention. Therefore, a prospective study would be one that specifically measures the sensitivity of Spanish OSIS. Equally of interest would be a systematic review including all OSIS validations to compile and subsequently carry out a comparative analysis of the psychometric properties in order to identify areas for future improvement and provide evidence of its validity and reliability in several contexts and populations.

Regarding the strengths, this study included a large sample, exceeding even the recommendations of authors such as Hair et al. [38] and the quality standards established by COSMIN RoB [32]. On the other hand, it addressed many psychometric properties making it a particularly comprehensive adaptation. Extraordinarily, its methodological quality was evaluated, which also became particularly high, providing a solid basis for its application in clinical and research contexts.

Conclusions

This study successfully performed the cross-cultural adaptation of the Oxford Shoulder Instability Score to Spanish and thoroughly evaluated the psychometric properties of this version, as well as its methodological quality. Its construct validity, internal consistency, test-retest reliability, measurement error, discriminant ability, ease and speed of completion were positively highlighted, as well as its high methodological quality. Therefore, this new version of the Oxford Shoulder Instability Score proved to be an ideal tool for the assessment of shoulder instability in Spanish-speaking subjects, which will favor its application in both clinical and research settings.

Abbreviations

Approx	approximate
COSMIN RoB	CONsensus-based Standards for the selection of health Measurement Instruments Risk of Bias checklist
e.g	for example
gl	degrees of freedom
ICC	Intraclass correlation coefficient
min	minutes
OSIS	Oxford Shoulder Instability Score
PROMs	Patient Reported Outcome Measures
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies
s	seconds
SDC	Smallest Detectable Change
sig	significance
SEM	Standard Error Measurement
SST	Simple Shoulder Test
WOSI	Western Ontario Shoulder Instability

Acknowledgements

The authors would like to thank the Research Group "Area of Physiotherapy CTS-305" of the University of Seville, Spain; for its contribution in this study.

Author contributions

RA and GC developed the conceptual framework and designed the study protocol. RA and MB conducted data acquisition and analysis. All authors: RA, GC, FE and MB; contributed to data interpretation. RA and GC drafted the manuscript. All authors reviewed and provided critical revisions to the drafts. All authors approved the final version of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Ethics Committee of the Andalusian Public Health System (No. 1267-N-21).

Consent for publication

All participants signed the informed consent form.

Clinical trial number

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Physiotherapy, Research Group "Area of Physiotherapy" CTS-305, University of Seville, Seville 41009, Spain

Received: 10 September 2024 / Accepted: 13 November 2024

Published online: 23 November 2024

References

- Bahk M, Keyurapan E, Tasaki A, Sauers EL, McFarland EG. Laxity testing of the shoulder: a review. *Am J Sports Med.* 2007;35:131–44.
- Juel NG, Natvig B. Shoulder diagnoses in secondary care, a one year cohort. *BMC Musculoskelet Disord.* 2014;15:89.
- Farber AJ, Castillo R, Clough M, Bahk M, McFarland EG. Clinical Assessment of three common tests for traumatic anterior shoulder instability. *J Bone Joint Surg Am.* 2006;88:1467–74.
- Whittle JH, Peters SE, Manzanero S, Duke PF. A systematic review of patient-reported outcome measures used in shoulder instability research. *J Shoulder Elb Surg.* 2020;29:381–91.
- Shields DW, Jefferies JG, Brooksbank AJ, Millar N, Jenkins PJ. Epidemiology of glenohumeral dislocation and subsequent instability in an urban population. *J Shoulder Elb Surg.* 2018;27:189–95.
- Lo IKY, Griffin S, Kirkley A. The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: the Western Ontario Osteoarthritis of the shoulder (WOOS) index. *Osteoarthr Cartil.* 2001;9:771–8.
- Mazzoni B, Cucchi D, Giovannelli T, Paci M, Arrigoni P, Nicoletti S. Translation, cross-cultural adaptation, and validation of the Italian version of the Oxford shoulder instability score. *Int Orthop.* 2019;43:2125–9.
- Van Kampen DA, Van Den Berg T, Van Der Woude HJ, Castelein RM, Terwee CB, Willems WJ. Diagnostic value of patient characteristics, history, and six clinical tests for traumatic anterior shoulder instability. *J Shoulder Elb Surg.* 2013;22:1310–9.
- Castagna A, Nordenson U, Garofalo R, Karlsson J. Minor shoulder instability. *Arthrosc J Arthrosc Relat Surg.* 2007;23:211–5.
- Brophy RH, Beauvais RL, Jones EC, Cordasco FA, Marx RG. Measurement of shoulder activity level. *Clin Orthop.* 2005;439:101–8.
- Wright RW, Baumgarten KM. Shoulder outcomes measures. *J Am Acad Orthop Surg.* 2010;18:436–44.
- Wormald JCR, Geoghegan L, Sierakowski K, Price A, Peters M, Jain A, et al. Site-specific patient-reported outcome measures for Hand conditions: systematic review of Development and Psychometric properties. *Plast Reconstr Surg-Glob OPEN.* 2019;7:e2256.
- Michener LA, Leggin BG. A review of self-report scales for the assessment of functional limitation and disability of the shoulder. *J Hand Ther.* 2001;14:68–76.
- Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability. The development and validation of a questionnaire. *J Bone Joint Surg Br.* 1999;81:420–6.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures: Spine. 2000;25:3186–91.
- van der Linde JA, van Kampen DA, van Beers LWAH, van Deuren 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.* 2015;10:146.
- Olyaei GR, Mousavi SJ, Montazeri A, Malmir K. Translation and validation study of the Persian Version of the Oxford shoulder instability score. *jmr.* 2016;10:24–8.
- Sonmezer E, Yosmaoglu HB, Doğan CD. The reliability and validity of the Turkish version of the oxford shoulder instability score. *Disabil Rehabil.* 2020;42:261–6.
- Hanna SS, Khaja AF, Bouhamra AK, Maqdes A. Validity and Reliability of the Oxford Shoulder Instability Score Translated into Arabic. *J Musculoskelet Disord Treat.* 2020;6.
- Skare Ø, Liavaag S, Reikerås O, Mowinckel P, Brox JI. Evaluation of Oxford instability shoulder score, Western Ontario shoulder instability index and Euroqol in patients with slap (superior labral anterior posterior) lesions or recurrent anterior dislocations of the shoulder. *BMC Res Notes.* 2013;6:273.
- Hernandez-Sanchez S, Hidalgo MD, Gomez A. Cross-cultural adaptation of VISA-P score for Patellar Tendinopathy in Spanish Population. *J Orthop Sports Phys Ther.* 2011;41:581–91.

22. Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? *Sci Transl Med*. 2016;8:341ps12.
23. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Qual Life Res*. 2021;30:2197–218.
24. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford shoulder score revisited. *Arch Orthop Trauma Surg*. 2009;129:119–23.
25. Yuguero M, Huguet J, Griffin S, Sirvent E, Marcano F, Balaguer M, et al. Adaptación transcultural, validación y valoración de las propiedades psicométricas, de la versión española del cuestionario western Ontario shoulder instability index. *Rev Esp Cir Ortop Traumatol*. 2016;60:335–45.
26. Membrilla-Mesa MD, Tejero-Fernández V, Cuesta-Vargas AI, Arroyo-Morales M. Validation and reliability of a Spanish version of simple shoulder test (SST-Sp). *Qual Life Res*. 2015;24:411–6.
27. Kirkley A, Griffin S, McIntock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). *Am J Sports Med*. 1998;26:764–72.
28. Godfrey J, Hamman R, Lowenstein S, Briggs K, Kocher M, Village G. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. *J Shoulder Elb Surg*. 2007;16:260–7.
29. Lippitt SB, Harryman DT, Matsen FA. A practical tool for evaluation of the function: the simple shoulder test. *Acad Orthop Surg*. 1993. <https://www.scienceopen.com/document?vid=2fd52a23-1132-4bf4-86b3-8e3066745a24> Accessed on 28 Aug 2024.
30. 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;21:808–14.
31. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1171–9.
32. Terwee CB, Bot SDM, De Boer MR, Van Der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42.
33. Irrgang JJ, Lubowitz JH. Measuring Arthroscopic Outcome Arthrosc. 2008;24:718–22.
34. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4:293–307.
35. Lachenbruch PA, Lwanga SK, Lemeshow S. Sample size determination in Health studies: a practical Manual. *J Am Stat Assoc*. 1991;86:1149.
36. Mishra P, Pandey C, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. *Ann Card Anaesth*. 2019;22:67.
37. Martínez-Ortega RM, Tuya-Pendás LC, Martínez-Ortega M, Pérez-Abreu A, Cánovas AM. El coeficiente de correlación de los rangos de spearman, caracterización. *Rev Haban Cienc Méd*. 2009;8.
38. Hair J, Anderson R, Tatham R, Black W. *Análisis Multivariante*. 5th ed. Madrid: Pearson Prentice Hall; 2004.
39. Goursuch RL. *Factor analysis*. 2nd ed. New York, NY: Lawrence Erlbaum Associates; 1983.
40. Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York, NY: McGraw Hill; 1994.
41. Fleiss JL. *The design and analysis of clinical experiments*. New York, NY: John Wiley & Sons, Inc.; 1986.
42. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–36.
43. Møller A, Bissenbakker KH, Arreskov AB, Brodersen J. Specific Measures of Quality of Life in patients with Multimorbidity in Primary Healthcare: a systematic review on patient-reported outcome measures' adequacy of measurement. *Patient Relat Outcome Meas*. 2020;11:1–10.
44. Rowe CR, Patel D, Southmayd WW. The Bankart procedure: a long-term end result study. *J Bone Jt Surg Am*. 1978;60:1–16.
45. Haley CCA. History and physical examination for shoulder instability. *Sports Med Arthrosc Rev*. 2017;25:150–5.
46. Basar S, Gunaydin G, Hazar Kanik Z, Sozlu U, Alkan ZB, Pala OO, et al. Western Ontario shoulder instability Index: cross-cultural adaptation and validation of the Turkish version. *Rheumatol Int*. 2017;37:1559–65.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.