J Shoulder Elbow Surg (2023) 32, 1348–1356

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

BASIC SCIENCE

The Spanish version of the Constant-Murley Shoulder Score: translation, cultural adaptation, and validity

Yaiza Lopiz, MD, PhD[a,b,*,1], Daniel Garríguez-Pérez, MD[a,1], Juan Pablo Scarano-Pereira, MD[a], Manuel E. Fuentes Ferrer, PhD[c], Camilla Arvinius, MD, PhD[a], Virginia Ponz, MD[a], Carlos García-Fernández, MD[a], Fernando Marco, MD, PhD[a,b]

[a]*Shoulder and Elbow Unit, Department of Traumatology and Orthopaedic Surgery, Clínico San Carlos Hospital, Madrid, Spain*
[b]*Department of Surgery, Complutense University, Madrid, Spain*
[c]*Clinical Trials and Research Unit, Preventive Medicine Department, Clínico San Carlos Hospital, IdISSC, Madrid, Spain*

Background: The Constant-Murley Score (CMS) is one of the most employed tools for assessing shoulder function. It was first devised in 1987 for the English population and is now widely used internationally. However, it had yet to be cross-culturally adapted and validated to Spanish, which is the world's second-most native language. Formal adaptation and validation of clinical scores is paramount for them to be used with rigorous scientific methodology.
Methods: Following international recommendations for the cross-cultural adaptation of self-report measures, the CMS was first adapted into Spanish in six stages: translation, synthesis, back-translation, a review by expert committee, pretesting, and final appraisal by expert committee. After conducting a pretest with 30 individuals, the Spanish version of the CMS was tested on 104 patients with various shoulder pathologies to assess content, construct, criterion validity, and reliability.
Results: No major conflicts were encountered in the process of cross-cultural adaptation, with 96.7% of pretested patients having a full understanding of every item in the test. The validation showed excellent content validity (content validity index = .90), construct validity (strong correlation between items within the same subsection of the test), and criterion validity (CMS - Simple Shoulder Test, Pearson r = .587, $P = .01$; CMS - American Shoulder and Elbow Surgeons, Pearson r = .690, $P = .01$). Reliability of the test was also excellent, with high internal consistency (Cronbach's $\alpha = .819$), interrater reliability (intraclass correlation coefficient = .982), and intrarater reliability (intraclass correlation coefficient = .937), without showing ceiling or floor effects.
Conclusion: Spanish version of the CMS has been proved to accurately reproduce the original score and to be easily comprehensible by native Spanish speakers with acceptable intrarater-interrater reliability and construct validity.

Introducción: La escala de Constant-Murley (CMS) es una de las más empleadas para evaluar la función del hombro. Se ideó por primera vez para la población inglesa en 1987 y actualmente es ampliamente empleada a nivel internacional. Sin embargo, su validación y adaptación transcultural no se han realizado al español, la segunda lengua nativa más hablada en el mundo. Actualmente no puede

The Ethics Committee for Clinical Research (CEIC) from Clínico San Carlos Hospital (Madrid, Spain) approved this study (internal code: 21/008-E).

[1] These authors contribute equally to this work.

*Reprint requests: Yaiza López, MD, PhD, Shoulder and Elbow Unit, Department of Traumatology and Orthopaedic Surgery, Clínico San Carlos, 5° Planta, Ala Sur. Calle Profesor Martín Lagos s/n 28004, Madrid, Spain.

E-mail address: yaizalopiz@gmail.com (Y. Lopiz).

https://doi.org/10.1016/j.jse.2023.01.032

resultar admisible el empleo de escalas en las que no tengamos la seguridad de que existe una equivalencia conceptual, cultural y lingüística entre la versión original y la empleada.
Material y Método: La versión traducida al español de la CMS se realizó siguiendo las recomendaciones internacionales: traducción, síntesis de la traducción, retrotraducción, revisión por comité de expertos, pretest y validación. Tras la realización del pretest en 30 individuos, la versión española de la escala de CMS se probó en 104 pacientes con diferentes patologías de hombro para evaluar las propiedades psicométricas de la escala: contenido, constructo, validez de criterio y fiabilidad.
Resultados: No se encontraron problemas importantes durante el proceso de adaptación transcultural con un entendimiento completo de todos los ítems del pretest por el 96.7% de los pacientes. La escala adaptada demostró una excelente validez de contenido (índice de validez de contenido = .90), de constructo (fuerte correlación entre ítems de la misma subsección del test), y de criterio (CMS-SST Pearson's r = .587, p = .01; CMS-ASES Pearson's r = .690, p = .01). La Fiabilidad del test resultó excelente, con una elevada consistencia interna (Cronbach's α = .819), fiabilidad interobservador (ICC = .982) e intraobservador (ICC = .937), sin efectos techo y suelo.
Conclusión: La versión Española de la CMS garantiza la equivalencia con respecto al cuestionario original. Los presentes resultados, sugieren que esta versión es válida, fiable y reproducible para la evaluación de la patología de hombro en nuestro entorno.
Level of evidence: Basic Science Study; Validation of Outcome Instrument

Keywords: Constant-Murley Score; Constant score; Spanish; validity; reliability; shoulder function

The infinite variability between individuals, the subjective way in which symptoms are expressed, and the challenge of measuring clinical signs imply that assessing a pathology or treatment could sometimes be more art than science. To mitigate this lack of accuracy in our measurements, we devise indexes, scores, scales, or other tools to quantify symptoms and signs in comparable figures.

Shoulder function is assessed by orthopedic surgeons using many clinical scores: the American Shoulder and Elbow Surgeons (ASES) score, Simple Shoulder Test (SST), Oxford Shoulder Score, and the Constant-Murley Score (CMS), among others. The latter has been one of the most employed shoulder function assessment tools since it was approved in 1989 by the Executive Committee of the European Society for Surgery of the Shoulder and Elbow. Although the CMS is widely used across the world, it has not been validated in every population and language (a simple translation of the questionnaire does not ensure its validity in another language or culture). This is the case for the Spanish language, the second-most spoken language in terms of the number of native speakers, and to the best of our knowledge, there has been no Spanish cross-cultural adaptation of the CMS. Therefore, this study aimed to evaluate the internal consistency, test-retest reliability, and construct validity of the translated and adapted Spanish version of the Constant score in patients with shoulder disorders.

Materials and methods

Study group

After obtaining approval from the institutional review board "ethics committee for clinical research" of our center (internal code: 21/008-E), participants were recruited in the outpatient clinic of a tertiary hospital in the city of Madrid, Spain, between February 2021 and March 2022. Patients over 18 years of age with ongoing shoulder pathology and able to speak and write Spanish fluently were included. Exclusion criteria were as follows: patients under 18 years of age, previous shoulder surgical procedure, moderate to severe cognitive impairment, peripheral or central nerve damage or diagnosed with nerve dysfunction, tumor, lack of sufficient knowledge of Spanish, and shoulder instability. We included shoulder instability as an exclusion criterion because of the inadequacy of the CMS in assessing this pathology has already been shown.[5]

Translation and cross-cultural adaptation

Translation and cross-cultural adaptation of the modified CMS[6] were performed according to the international guidelines recommended by Beaton et al[2] This guideline includes an adaptation process of six phases: translation, synthesis, back-translation, review by expert committee, pretesting, and final appraisal by expert committee.

First, the CMS was independently translated by two bilingual individuals: translation 1 was conducted by an orthopedic surgeon and translation 2 was performed by a professional writer without previous medical knowledge. A synthesis of both versions was made, and discrepancies were analyzed in a meeting with these translators. The resulting translation was reviewed by the whole team and checked for any changes needed for cultural adaptation.

Back-translation into English was then performed independently by two professional bilingual translators who had never seen the original CMS. Thus, back-translation 1 and 2 were generated, synthesized in one single version, and reviewed by the team for cultural mismatch. The aim of this new translation was to ensure that the Spanish version faithfully reflected the content and meaning of the original English version of the CMS.

The next step was the drafting of a fully culturally adapted version considering all existing versions obtained from the previous steps through a committee meeting of multidisciplinary experts. In addition, the author of the original questionnaire, Dr. Constant, was contacted to inform him of the process.

Once the preliminary version of the scale had been drafted, a pretest evaluation was carried out on a sample of 30 subjects with painful shoulder (15 women and 15 men) to determine the acceptability and comprehensibility of the translation as had been done in previous studies.[13]

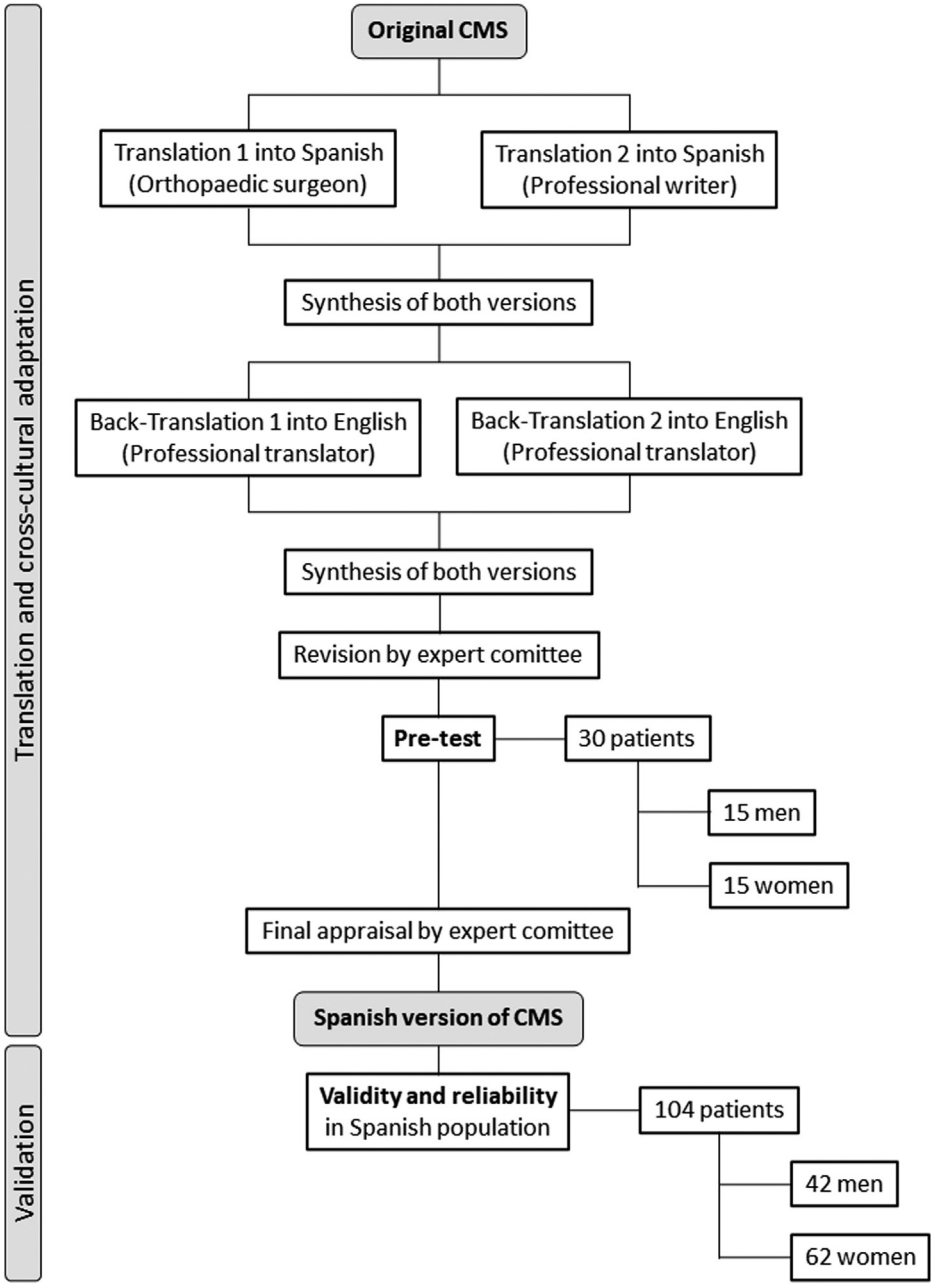

Figure 1 Procedures for cross-cultural adaptation of the Constant-Murley Score (CMS).

The preliminary results were reviewed again by a committee of experts and the necessary changes and improvements were included to produce the final version of the Spanish Constant-Murley Scale (ES-CMS). Patients from the pilot study were not included in the subsequent validation study. Figure 1 summarizes the whole procedure of translation and cultural adaptation.

Validation of the Spanish version of the CMS (ES-CMS)

Data collection for the validation process of the score was performed by evaluating each patient 3 times: the first (ES-CMS1) and the second (ES-CMS2) assessments were made on the same day by two different examiners; the third (ES-CMS3) assessment was made 2 weeks after the first two by the same examiner that had assessed ES-CMS1. Patients were instructed not to take treatment during these 2 weeks.

For sample size calculation, 10 subjects were needed for each item of the score, giving a minimum total of 80 patients for the 8 items of the score. In similar studies, losses have been as high as 30% of patients, so a final sample size of 104 patients was estimated to compensate for possible losses.

Table I Baseline characteristics and functional results of the included sample (n = 104)

Age	59.25 ± 12.75
Sex, male/female	42 (40.4)/62 (59.6)
Shoulder, right/left	57 (54.8)/47 (45.2)
Handedness, right/left	95 (91.3)/9 (8.7)
Diabetes	16 (15.4%)
Hypothyroidism	12 (11.5%)
Occupation	
Manual labor	35 (33.7)
Office	24 (23.1)
Housemaker	5 (4.8)
Retired or pensioner	34 (32.7)
Unemployed or student	5 (4.8)
N/A	1 (1)
Educational stage	
Primary	28 (26.9)
Secondary	45 (43.3)
Higher education	30 (28.8)
N/A	1 (1)
Functional scores*	
ASES	43.75 ± 18.66
SST	44.87 ± 23.97
CMS	57.81 ± 17.40

ASES, American Shoulder and Elbow Surgeons; *SST*, Simple Shoulder Test; *CMS*, Constant-Murley Score; *N/A*, not available.
Data are presented as "number of patients (percentage)" for categorical variables and "mean ± standard deviation" for quantitative variables.
* Data are presented as mean ± standard deviation.

Data analysis and psychometric scale properties

Content validity was measured by examination of the data distribution shape, alongside ceiling and floor effects. Construct validity was assessed using exploratory factor analysis performed by principal component analysis. Prior to this, a Kaiser-Meter-Olkin test was performed to evaluate how suitable our data was to factor analysis, and the Bartlett's Test of Sphericity was used to check for redundancies between the items. A rotated component matrix was generated with Varimax rotation with Kaiser normalization to assess which items are associated with which factor and how strong their correlation is (Pearson correlation coefficient). Criterion validity was determined using the concurrent validity method by evaluating the relationships between the CMS and the results of the SST and the ASES Standardized Shoulder Assessment Form, which have both been validated for the Spanish population,[10,13] calculating the Pearson correlation coefficient between them. The patient was given these two scores at the same time as the ES-CMS1.

Reliability of the ES-CMS was measured by analyzing internal consistency and interrater and intrarater reliability. Internal consistency was measured using Cronbach's alpha, with closest values to 1 meaning high internal consistency and closest values to 0, low internal consistency. Interrater/intrarater reliability was analyzed using the intraclass correlation coefficient (ICC) between ES-CMS1 and ES-CMS2 for interrater reliability and between ES-CMS1 and ES-CMS3 for intrarater reliability. Bland-Altman plots were also generated to depict possible differences between measures.

Statistical analysis was performed with IBM SPSS software (version 22; IBM, Armonk, NY, USA). *P* values of .05 or less were considered significant.

Results

Translation and cross-cultural adaptation

The ES-CMS can be viewed in Supplementary Appendix S1. Different items from the original questionnaire were adjusted to adapt the modified CMS to the Spanish language and culture. Most notably, questions regarding activities of daily living and recreational activities were changed to be more idiomatic in our language: "how much of your work/activity does your shoulder allow" was adapted to something like "how many of your activities can you perform in spite of your shoulder pain". All questions and instructions addressed to patients were written using a formal imperative and the third person singular, verb tenses that do not exist in English.

Demographics

One hundred four patients were included in the validation process. No patients were lost to follow-up. Baseline characteristics of the sample and the mean scores of the three questionnaires employed (ASES, SST, and ES-CMS) are given in Table I.

Content validity

After revision of the final ES-CMS, all members of the experts committee agreed that every item of the questionnaire adequately represented the attributes and dimensions affected by a painful shoulder and its functionality. Moreover, there were no floor and ceiling effects when more than 15% of the respondents achieved the highest or lowest possible scores.

Construct validity

The Kaiser-Meter-Olkin index was .789, indicating good sampling adequacy to perform a factor analysis. Bartlett's Test of Sphericity was statically significant ($P < .001$), proving the null hypothesis wrong of the CMS items not being correlated and, thus, allowing a factor analysis.

Principal component analysis concluded that the ideal number of factors representing the dataset was 3, with the lowest Eigen value above 1 (Fig. 2). Finally, a rotated component matrix was generated (Table II), showing that items related to mobility were associated with factor 1, items related to daily activities were associated with factor

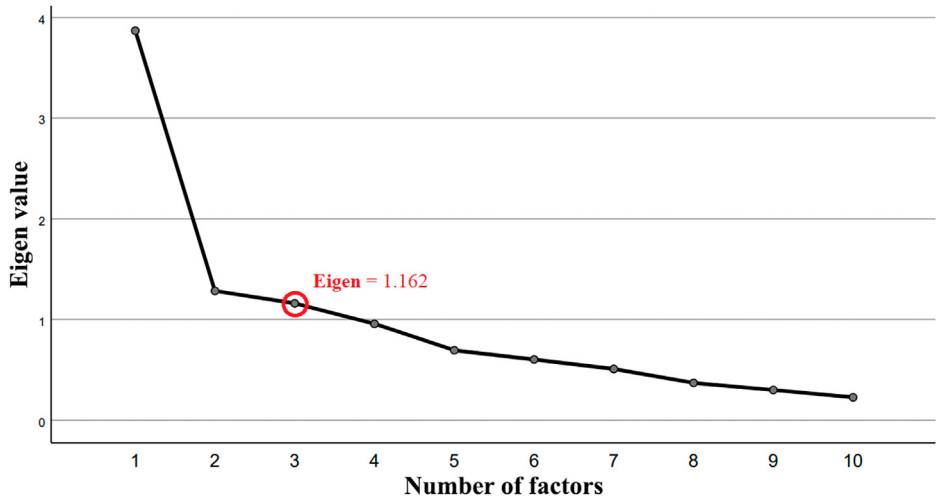

Figure 2 Screen plot showing the results of the principal component analysis in which a 3-factor model grouped the items in the CMS with the lowest Eigen value above 1, thus representing the most appropriate model. *CMS*, Constant-Murley Score.

2, and pain was associated with factor 3. Sleep disturbance and the level to which the arm can be elevated comfortably were the only two items that were not clearly associated with any of the generated three factors. Table III represents the generated correlation matrix of every item in the CMS. The strongest correlations were found between flexion and abduction ($P = .684$, $P = .01$) and between flexion and external rotation ($P = .707$, $P = .01$).

Criterion validity

Pearson correlation coefficient was .587 ($P = .01$) between the ES-CMS and SST and .690 ($P = .01$) between the ES-CMS and ASES (Fig. 3).

Internal consistency

Cronbach's alpha was .819, corresponding to an adequate internal consistency for the score. This value diminished each time that any item of the score was removed from the analysis, demonstrating a positive contribution of every item in terms of internal consistency.

Interrater reliability

There was a statistically significant correlation between ES-CMS1 and ES-CMS2 ($P = .983$, $P = .01$). The ICC between ES-CMS1 and ES-CMS2 was .982, and the differences between the means of both variables were not significantly different when assessed with the Student t-test ($P = .068$). A Bland-Altman plot was generated, showing overall differences of less than 10 points between ES-CMS1 and ES-CMS2 (Fig. 4).

Table II Rotated component matrix

Items	Factor 1	Factor 2	Factor 3
Pain	0.234	0.293	**0.778**
Daily activities	0.203	**0.820**	0.023
Recreational activities	0.110	**0.838**	0.010
Sleep	0.347	−0.180	0.193
Comfortable arm height	0.382	0.291	−0.643
Flexion	**0.801**	0.324	−0.188
Abduction	**0.760**	0.253	−0.151
External rotation	**0.764**	0.314	−0.103
Internal rotation	**0.720**	−0.108	0.296
Strength	**0.708**	0.217	0.060

Values in the table express the correlation index between each item and the factors generated via principal component analysis. Strongest correlations appear in bold text, showing how each variable belongs in a certain factor, except for sleep and comfortable arm height, which did not show strong correlations with any of the generated factors.

Intrarater reliability

There was a statistically significant correlation between ES-CMS1 and ES-CMS3 ($P = .937$, $P = .01$). The ICC between ES-CMS1 and ES-CMS3 was .934. Here, the differences between the means of both variables were significantly different when assessed with the Student t-test ($P = .022$). A Bland-Altman plot was generated, showing overall differences of less than 10 points between ES-CMS1 and ES-CMS3 (Fig. 4).

Discussion

Although the Spanish Society of Shoulder and Elbow Surgery (Sociedad Española de Cirugía de Hombro y Codo) provided a translated version of the Constant-Murley Score for use by Spanish-speaking orthopedic surgeons (https://

Table III Correlation matrix between every item of the CMS

	Pain	Daily activities	Recreational activities	Sleep	Comfortable arm height	Flexion	Abduction	External rotation	Internal rotation	Strength
Pain	1	-	-	-	-	-	-	-	-	-
Daily activities	.192	1	-	-	-	-	-	-	-	-
Recreational activities	.192	.543†	1	-	-	-	-	-	-	-
Sleep	.075	.038	.044	1	-	-	-	-	-	-
Comfortable arm height	−.109	.244*	.260†	.044	1	-	-	-	-	-
Flexion	.177	.398†	.317†	.086	.467†	1	-	-	-	-
Abduction	.145	.293†	.330†	.209*	.360†	.684†	1	-	-	-
External rotation	.137	.418†	.324†	.086	.336†	.707†	.578†	1	-	-
Internal rotation	.263†	.102	.088	.185	0.087	.455†	.337†	.504†	1	-
Strength	.278†	.335†	.159	.095	.264†	.553†	.566†	.521†	.339†	1

CMS, Constant-Murley Score.
Correlation values are expressed using Pearson correlation coefficients.
* Statistical significance of $P < .05$.
† Statistical significance of $P < .01$.

sechc.es/test-score-hombro), to the best of our knowledge, no Spanish validation or cross-cultural adaptation has been published to date. The validation process for a scale involves more than simply translating the questionnaire. Moreover, cross-cultural adaptation does not assume that the psychometric properties of the adapted scale are the same as those of the original scale; therefore, it is essential to evaluate the psychometric properties of the adapted version. Given the prevalence and socioeconomic impact of shoulder disorders, we believe that a Spanish cultural adaptation and validation of the CMS would be extremely beneficial for Spanish-speaking surgeons and patients.

The strength of the present study is that the translation procedure followed the international guidelines established by the American Academy of Orthopaedic Surgeons (AAOS) Outcome Committee, leading to more explicit and reliable results.

In relation to the process of translation and cross-cultural adaptation, there was controversy regarding some of the questions of the score. The term ''activities of daily living'' is well-accepted in Spanish medical literature, but Spanish speakers without a medical background can find this term confusing. After discussing various options, we ended up choosing ''daily tasks'' as the easiest term to understand in our language. We also discussed whether these questions related to activities of daily living should be asked as ''how does your shoulder affect/allow…'' or ''how does your shoulder pain affect/allow…'', and finally we decided that asking directly how pain affects this dimension could be easier for patients to understand.

The study by Blonna et al[3] demonstrated that the reliability of the CMS could be significantly improved in the hands of a rater, experienced and inexperienced, if a standardized test protocol were used. To achieve this, we included a full-page protocol which gives instructions to the examiner on how to gather data on each of the items of the score, focusing on movement and strength, which can be the most variable items between examiners since they are based on objective measures that can have great inter-rater variability if not performed in a homogeneous way. These instructions included explanatory 3-dimensional figures to show how to measure external rotation and strength, which we found to be the anatomical positions that raised the most queries when explained only with text instructions. Both the ES-CMS and its protocol are provided together in Supplementary Appendix S1 and are designed to be printed on both sides of a sheet.

The modified CMS advises that strength should be measured by either an Isobex (CURSOR AG, Bern, Switzerland) device or a defined spring balance technique. This measurement is recorded at 90° of abduction in the scapular plane. The wrist is in a position of pronation, so that the hand is facing downward. We realized during the adaptation process that the term ''in the plane of the scapula'' is not easily understood by examiners, so we have introduced an explanatory note to better describe this position, adding an indication to position the arm in 90° of abduction and bring it forward about 30°. We also believe that the fixed devices that were employed to measure strength when devising this score are now somewhat obsolete. For the validation of the ES-CMS, a handheld dynamometer (Lafayette MMT, Manual Muscle Tester; Lafayette Instrument, Lafayette, IN, USA) was used instead, since portable devices have proved valid and reliable to measure shoulder abduction strength[7] and are much more convenient for everyday use in clinics. Another important aspect of measuring shoulder strength in our environment is the unit of measurement. The original CMS measures strength in pounds, and the result is directly converted to points on the score, whereas we measure it in kilograms and therefore need to convert it to calculate the score: the kilograms of shoulder strength obtained on the

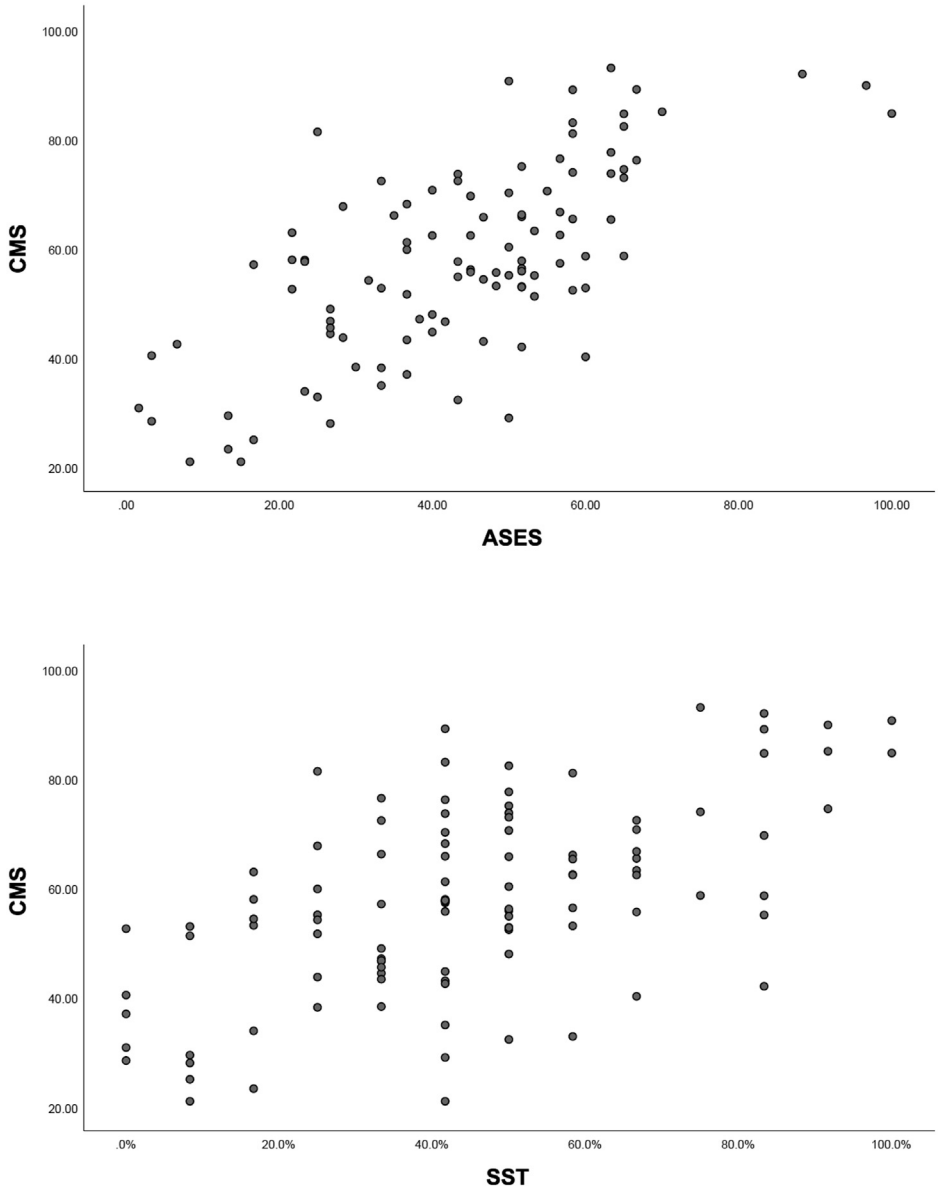

Figure 3 Scatter plots showing a statistically significant moderate correlation between the CMS and other functional scores (ASES and SST, respectively). *CMS*, Constant-Murley Score; *ASES*, American Shoulder and Elbow Surgeons; *SST*, Simple Shoulder Test.

dynamometer are multiplied by 2.2 to convert them to pounds, that is, to points.

In terms of psychometric properties, reliability (ability of the scale to consistently and accurately measure the characteristic it intends to measure), can be determined in several ways: 1) as constancy or stability over time, which refers to the ability of the scale to obtain the same measurements in the same subjects at different times (test/retest: in the present study ES-CMS1/ES-CMS3 results) by the same examiner (intrareliability) or different examiners (interreliability); 2) confidence as equivalence of results obtained by the same subjects with two different questionnaires (in the present study ES-CMS1, SST, and ASES); and 3) reliability as internal consistency in reference to the degree of homogeneity of the scale questions. Internal consistency is calculated using Cronbach's alpha and is a measure based on the correlations between different items in the same test (or the same subscale in a larger test). This parameter represents the degree to which every test item proposed to assess shoulder function produces similar scores. The analysis showed that the ES-CMS presented high internal consistency (Cronbach's alpha = .819) and excellent psychometric properties in interobserver and intraobserver reliability (ICC of .982 and .934, respectively, considered excellent above .75),[1] without floor or ceiling effects. The interobserver reliability was higher than the intraobserver reliability and presented fewer differences, possibly because ES-CMS1 and ES-CMS2 were performed in less than 1 hour, which reduced the risk that more subjective responses, such as

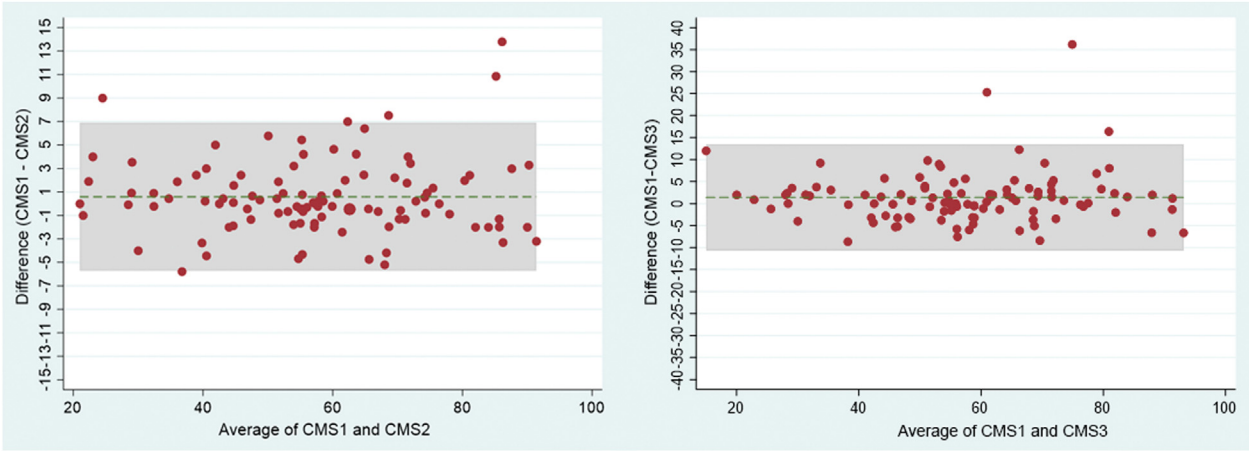

Figure 4 Bland-Altman plots showing the differences in scores between CMS1 and CMS2 (interrater) and between CMS1 and CMS3 (intrarater). *CMS*, Constant-Murley Score.

pain, would undergo significant changes. Despite this, the intraobserver reliability continued to be higher than that found in the Chinese[14] (ICC = .827) and Turkish[4] (ICC = .86) translations and similar to that obtained in the Danish[11] (ICC = .93-95) and Greek CMS[12] (ICC = .95).

Another psychometric property of a test is its validity. Validity is the ability of a scale to measure what it was designed to measure. There are different types of validity of a clinical questionnaire. A) Construct validity is the most general term and can encompass all other types of validity, referring to the ability of the instrument to meet the measurement hypothesis for which it is designed. For its assessment, we conducted a factorial analysis of the elements that make up the scale. This analysis would answer the question: can the scale be grouped into different factors of related variables that represent the dimensions of the evaluated construct? To do this, we grouped it into three factors. The results obtained showed that all the elements of the scale were adequately represented in the factors except the items ''sleep disturbance'' and ''arm elevation in daily activities''. These two items have low correlations with other items on the scale and the lowest item-total connection. This may be because both items were too subjective and variable over time. Although these two items provide the least consistency to the questionnaire, we do not believe that both items are degrading for the scale, and the data obtained would not justify their exclusion from the measurement instrument. B) Content validity describes the suitability of the questionnaire items to measure what is intended to be measured. This validity is based on the criteria of the research team since there are no objective methods that guarantee the suitability of the questions. C) Criterion validity examines whether a test ''measures what it claims'' and is evaluated in relation to other tests already validated to serve the same purpose. In this study, we employed two scales widely used in shoulder function evaluation: the SST and ASES. The values obtained were .587 and .690. Between .81 and 1.0 is considered excellent, between .61 and .80 very good, between .41 and .60 good, between .21 and .40 fair, and between 0 and .20 poor.[9] Therefore, the correlation turned out to be higher with the ASES questionnaire. In the Arabic and Greek adaptations,[8,12] the disabilities of the arm shoulder and hand questionnaire (DASH) and Quick-DASH were used to analyze this parameter, finding a strong correlation (.820 and .84, respectively) between both scales. For the Chinese adaptation,[14] the visual analog scale and the short form-36 health survey (SF-36) were used, demonstrating an adequate correlation with the visual analog scale (r = .497) but not with the SF-36 (r = .135). Differences in contrast validity with respect to the ES-CMS can be explained by the inherent structural difference of each of the questionnaires employed. None of the other CMS translations assessed criterion validity using the SST.

Regarding the use of our validated tool in other Latin American countries, it may seem controversial based on the different evolution of the Spanish language throughout them. We believe however that cultural identity prevails over strict wording differences. Furthermore, the items evaluated in the CMS are simple issues centered on pain and activities of daily living which are readily understood in every Latin American country in plain Spanish (Castilian). Nevertheless, even if the transcultural adaptation of the scale may be understandable and valid for Latin America, this does not mean that psychometric properties can be assumed as equal to Spain. That is why we consider that the evaluation of the aforementioned properties is mandatory, and the tool hereby described could be a starting point for it.

Study limitations

The present study has several limitations. First, the absence of an exhaustive sensitivity analysis, which would involve more factors such as the effectiveness of the

intervention as well as the type and severity of the disease. Second, only patients with unilateral shoulder pain were studied, which does not accurately represent the population with shoulder discomfort, where bilateral involvement is not uncommon. Third, the time between each evaluation was two weeks to prevent the patients from remembering the data provided on the first day. However, this period could result in a variation in the final outcome as a consequence of the modification of some of the items such as pain; although to minimize this variation, patients were asked not to undergo any treatment during those two weeks.

Conclusions

The process of adapting the CMS to the Spanish culture and language through translation-back translation guarantees conceptual, cultural, and linguistic equivalence with respect to the original questionnaire. The adapted ES-CMS showed excellent results in terms of reliability, internal consistency, and construct validity and is useful in the evaluation of shoulder disorders for Spanish patients and doctors. The factorial analysis revealed that the items such as sleep and the ability to raise the hand during daily activities were not informative items, but neither can they be considered degrading items for the scale.

Disclaimers:

Funding: No funding was disclosed by the authors.
Conflicts of interest: The authors, their immediate families, and any research foundation with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jse.2023.01.032.

References

1. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002;11:193-205. https://doi.org/10.1023/a:1015291021312
2. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000;25:3186-91.
3. Blonna D, Scelsi M, Marini E, Bellato E, Tellini A, Rossi R, et al. Can we improve the reliability of the Constant-Murley score? J Shoulder Elbow Surg 2012;21:4-12. https://doi.org/10.1016/j.jse.2011.07.014
4. Celik D. Turkish version of the modified Constant-Murley score and standardized test protocol: reliability and validity. Acta Orthop Traumatol Turc 2016;50:69-75. https://doi.org/10.3944/AOTT.2016.14.0354
5. Conboy VB, Morris RW, Kiss J, Carr AJ. An evaluation of the Constant-Murley shoulder assessment. J Bone Joint Surg Br 1996;78:229-32.
6. Constant CR, Gerber C, Emery RJH, Søjbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. J Shoulder Elbow Surg 2008;17:355-61. https://doi.org/10.1016/j.jse.2007.06.022
7. Karabay D, Yesilyaprak SS, Sahiner Picak G. Reliability and validity of eccentric strength measurement of the shoulder abductor muscles using a hand-held dynamometer. Phys Ther Sport 2020;43:52-7. https://doi.org/10.1016/j.ptsp.2020.02.002
8. Maqdes A, Hanna SS, Bouhamra AK, Khaja AF. Cross-cultural adaptation and translation of the Constant Murley score into Arabic. SICOT J 2020;6:44. https://doi.org/10.1051/sicotj/2020042
9. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol 2003;56:730-5. https://doi.org/10.1016/s0895-4356(03)00084-2
10. 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. https://doi.org/10.1007/s11136-014-0760-2
11. Moeller AD, Thorsen RR, Torabi TP, Bjoerkman A-SD, Christensen EH, Maribo T, et al. The Danish version of the modified Constant-Murley shoulder score: reliability, agreement, and construct validity. J Orthop Sports Phys Ther 2014;44:336-40. https://doi.org/10.2519/jospt.2014.5008
12. Ntourantonis D, Panagopoulos A, Iliopoulos I, Tatani I, Tsoumpos P, Kouzelis A, et al. Translation, cultural adaptation, validity and reliability of the Greek version of the modified Constant Score. JSES Open Access 2017;1:45-50. https://doi.org/10.1016/j.jses.2017.02.004
13. Vrotsou K, Cuéllar R, Silió F, Rodriguez MÁ, Garay D, Busto G, et al. Patient self-report section of the ASES questionnaire: a Spanish validation study using classical test theory and the Rasch model. Health Qual Life Outcomes 2016;14:147. https://doi.org/10.1186/s12955-016-0552-1
14. Yao M, Yang L, Cao Z, Cheng S, Tian S, Sun Y, et al. Chinese version of the Constant-Murley questionnaire for shoulder pain and disability: a reliability and validation study. Health Qual Life Outcomes 2017;15:178. https://doi.org/10.1186/s12955-017-0752-3