Transcultural reliability and validity of an Italian language version of the Constant–Murley Score

Matteo Carosi¹, Giovanni Galeoto², Stefano Di Gennaro³, Anna Berardi², Donatella Valente² and Annamaria Servadio¹, ⁴

Abstract

Purpose: The present study was designed to carry out an Italian translation of the Constant–Murley Score (CMS-IT) and, subsequently, evaluate its reliability and validity. Methods: This study included adults diagnosed with any type of clinical shoulder dysfunction who could read and respond to the questionnaires. Those individuals who underwent surgeries of any kind on the affected shoulder during the previous 12 months and individuals with shoulder instabilities were excluded. All of the participants were evaluated by two operators, and the CMS-IT, Disabilities of the Arm, Shoulder, and Hand (DASH) questionnaire, and visual analog scale (VAS) were administered. The internal consistency was evaluated using Cronbach’s α, whereas the intra-rater and inter-rater reliabilities were evaluated using the intraclass correlation coefficient (ICC). The validity of the construction was evaluated using Pearson’s correlation coefficient between the scores of the administered scales. Results: A population of 72 individuals participated in this study. The internal consistency of the CMS-IT showed a value of 0.81. The ICC values showed that inter- and intra-rater reliability were 0.994 and 0.963, respectively. CMS-IT is inversely correlated with the VAS (−0.55) and DASH (−0.47) scales. Conclusions: This study revealed that the CMS-IT contained good internal consistency and good reliability. The results suggested that the CMS-IT questionnaire is a reliable and valid tool for assessing the shoulder dysfunctions of the Italian population, and it deserves broad applications in both clinical practice and research contexts. The scale can also be used as an alternative to the current “gold standard” VAS and DASH.

Keywords

Constant–Murley scale, reliability, shoulder evaluation, shoulder pain, validation study

Introduction

Shoulder pain accounts for an expressive prevalence of consultations with general practitioners and orthopedic surgeons. These individuals often present with various complaints, such as mobility deficits and pain, which directly affect their upper limb function. To perform as comprehensive a clinical assessment as possible, it has been recommended that the individual be assessed using instruments that allow inferences regarding function.¹ Function is a construct, which is a latent variable that cannot be directly observed. Therefore, the utilization of functional scores is an appropriate option for measuring it. A review of the literature shows that there are many evaluation tools,² and that there is no “gold standard instrument” for evaluating...
the outcomes of shoulder function in the general population. Among the different scales available for the evaluation of the upper limb, the Constant–Murley Score (CMS) is one of the most commonly used instruments. This assessment scale requires only few minutes to be administered, and it is very intuitive and easy to understand both for the operators and patients. This minimizes operator-dependent errors, allows the patients to respond correctly and effectively, and provides the health professionals with reliable feedback while collecting the data in the most accurate and standardized way. Practitioners are also placed in a position to shape a more effective rehabilitation plan aimed at the recovery of the patient, while eliminating general rehabilitation projects for all patients and customizing individual rehabilitation plans.

The CMS was presented in 1987 as an instrument for evaluating the overall shoulder function, irrespective of the diagnosis. It was approved and recommended by the executive committee of the Shoulder and the Elbow, and it has been widely used as an assessment method ever since. The CMS was created between 1981 and 1986 by Christopher Constant and Alan Murley, while the scale itself was published in 1987. However, it has been largely criticized over the years for its lack of methodology. In an attempt to clarify certain aspects related to the administration of the CMS, in 2008, the original authors published an article with modifications and guidelines for the instrument’s use. A visual analog scale (VAS) was suggested for the pain item and some of the activity of daily living (ADL) questions, and specific instructions were provided on how to evaluate the strength component. It was also stated that the CMS was not valid for evaluating episodic severe pain, such as that experienced in a dislocation. According to one recent systematic review, five individual pathologies (i.e. subacromial, fractures, arthritis, instability, and frozen shoulder) and two additional groups (i.e. various pathologies and healthy subjects) were considered in the CMS. However, the results of that review suggested that the current evidence supported the CMS as a “gold standard instrument” for the subacromial pathologies only, while the data were inconclusive for other shoulder conditions. The authors suggested further prospective studies exploring the psychometric properties of the scale, particularly for fractures, arthritis, instability, and frozen shoulders.

Recently, the CMS was validated in several languages: Danish, Brazilian Portuguese, Turkish, Chinese, and Greek. Moreover, from 2008 to 2018, more than 70 studies were conducted in Italy in which the CMS was used but not validated. The CMS has also been used as the “gold standard instrument” in certain studies, such as one conducted in Italian for the validation of the Oxford Shoulder Score, even though the CMS has not yet been validated in Italian.

The purpose of this study was to translate, culturally adapt, and validate the CMS in a sample of Italian individuals with shoulder disorders using a cross-sectional study.

Methods
This study was conducted by a research group composed of medical doctors and rehabilitation professionals from the Sapienza University of Rome and the Rehabilitation and Outcome Measures Assessment (ROMA) association. During the last few years, the ROMA association has addressed the validation of many outcome measures in Italy.

Instruments
All of the participants enrolled in this study received an informed consent form describing the type of study that was being carried out and what the evaluation was based on, as well as a request for authorization to use the data for the purposes of this study. The second part contained a questionnaire for the collection of anamnestic data, which is useful for investigating the possible cause of the pain. For example, it contained questions about which was the dominant hand, the participant’s occupation, sports activities (if at a competitive level), and whether the participant’s sleep was disturbed or not. In the last part of the document, the three evaluation scales were reported: the CMS, the “gold standard instrument” VAS, and Disabilities of the Arm, Shoulder, and Hand (DASH).

The CMS evaluates pain and disability based on both subjective and objective parameters. It has a total score of 100 points, with 35 assigned to the subjective parameters and 65 to the objective parameters; the higher the score, the greater the quality of the function. The subjective component is made up of five items that evaluate the pain and ADLs; however, the objective component measures the active range of motion and strength. The subjective module evaluates the limitations that the patient experiences regarding work and recreational activities. The pain is assessed using a VAS of 15 cm.

The DASH is a self-administered region-specific outcome instrument developed as a measure of self-rated upper-extremity disability and symptoms.

Translation and cultural adaptation
After receiving consent from the developers of the instrument, the CSM was translated from English to Italian using the Translation and Cultural Adaptation of Patient Reported Outcomes Measures—Principles of Good Practice guidelines. The first phase of the study involved the translation from English into Italian by two native English speakers not connected to each other, thus producing two Italian versions. An independent native speaker of the target language, with Italian as second language, who was not involved in any of the forward translations synthesized the results of the translations. Working from the temporary version of the questionnaire, three Italian translators then translated the questionnaire back into the original language without having seen the original version. The back-translated version of the instrument was then compared with the original. To adapt
the translated version to the Italian culture, two Italian rehabilitation professionals who were familiar with both English and Italian reviewed the translated version, and then, they reworded and reformulated certain items to minimize any differences from the original version.

Sample and validation procedures

The final version of the questionnaire (CMS-IT) was administered to a sample of participants recruited from two Italian clinics. The individuals were informed about the procedures and aims of the study, and those who were interested in participating were screened based on the inclusion and exclusion criteria. Based on an analysis of the other CMS validation studies, the following criteria were formulated: ≥18 years old, able to read and respond to the questionnaire, and diagnosed with any clinical shoulder dysfunction. Individuals with cognitive diseases or damage to the peripheral or central nerves, non-Italian speakers, patients that underwent surgeries of any kind on the affected shoulder during the previous 12 months, and patients with shoulder instabilities were not included in this study.29

The CMS-IT, VAS, and DASH were administered for the first time (T₀) during the prehospitalization period or during an outpatient visit by a physiotherapist (rater 1), which was the same throughout the duration of the study. A subgroup of the population was reevaluated to examine the intra-rater reliability, and rater 1 administered the questionnaires twice to the same patient 4–7 days apart. Again, the subgroup of the population was reevaluated to examine the inter-rater reliability, and an orthopedic surgeon (rater 2) administered the questionnaires to the same participants that rater 1 did at T₀.

Data analysis

Consistent with the Consensus-based Standards for the Selection of Health Measurement Instruments checklist recommendations,30 the reliability and validity of the CSM-IT were assessed as follows. The Cronbach’s α for the internal consistency needed to be >0.7 to establish the degree of agreement between the various items. The intra-class correlation coefficient (ICC) of intra-rater and inter-rater reliabilities needed to be >0.7 to establish the stability of the individual measurements carried out at different times. Pearson’s correlation coefficient between the CSM-IT and the Italian versions of the DASH and the VAS needed to be >0.5 or <−0.5, where the positive values indicated a positive linear correlation and the negative values indicated a negative linear correlation. The significance level was set for a p value of less than or equal to 0.05. All of the statistical analyses were performed using IBM SPSS Statistics for Windows (Version 23.0; IBM Corp., Armonk, New York, USA).

Results

In July 2018, following the international guidelines,28 the final culturally adapted CMS-IT8 was formed. The cultural adaptation process resulted in the modification of item 2, which was related to ADL, in particular, professional skills (item 2.B), recreational activities (item 2.C), and the analysis of the height at which the patient could raise his/her hand (item 2.D). The aim of this approach was to make the language more understandable for both the interviewer and interviewee.

From September 2018 to November 2018, 88 individuals from two Italian clinics were examined, and of these, 16 were excluded from the study because they did not want to participate. Therefore, 72 individuals agreed to participate, and they completed the study. The demographic characteristic of the study sample are reported in Table 1. The mean ± standard deviation of the CMS-IT score was 51.15 ± 19.80. Cronbach’s α showed a value of 0.81, which indicated that a strong correlation was observed in each item-total correlation. As presented in Table 2, the α did not significantly increase with the hypothetical exclusion of any item, confirming the score arrangement of the items. Thirty of the participants

| Characteristic                  | Population = 72 | Characteristic                  | Population = 72 |
|--------------------------------|-----------------|--------------------------------|-----------------|
| Gender: men, n (%)             | 37 (51.4)       | Play an instrument, n (%)       | 2 (2.8)         |
| Age, mean (SD)                 | 60.6 (11.46)    | Guitar                          | No              |
| Employment, n (%)              |                 | No                             | 70 (97.2)       |
| Housewives                     | 34 (47.3)       | Dominant shoulder affected, n (%)| 52 (72)         |
| Employee                       | 12 (16.7)       | Diagnosis, n (%)                |                 |
| Driver                         | 3 (4.2)         | RCT                            | 54 (75)         |
| Cook                           | 3 (4.2)         | SMI + BLCI                      | 8 (11.1)        |
| Other                          | 20 (27.6)       | CT                             | 3 (4.2)         |
| Sport, n (%)                   |                 | SLAP                           | 2 (2.8)         |
| Walk                           | 6 (8.3)         | Arthrois                       | 2 (2.8)         |
| Gym                            | 4 (5.6)         | AD                             | 2 (2.8)         |
| Swimming                       | 3 (4.2)         | AC                             | 1 (1.4)         |
| Postural gymnastics            | 2 (2.8)         |                                |                 |
| Other                          | 13 (18.0)       |                                |                 |
| No                             | 44 (61.1)       |                                |                 |

Table 1. Demographic characteristic of the 72 participants.
(41.67%) underwent procedures to test the reliability of the administration of rater 1 and rater 2 through intra-operator and inter-operator test–retest methods. In Tables 3 and 4, the ICCs of the individual items are presented. The means of the ICCs with regard to the inter-rater and intra-rater reliabilities were 0.994 (range = 0.987 to 0.997, \( p < 0.001 \)) and 0.963 (range = 0.926 to 0.982, \( p < 0.001 \)), respectively. Finally, results showed that the CMS-IT was inversely correlated with the VAS (−0.55) and DASH (−0.47) scales, which means that as the CMS-IT score increased, the scores of the other scales decreased.

### Discussion

The CMS is a widely used tool that is recommended by the European Society for Surgery of the Shoulder and Elbow. Roy et al.\(^{31} \) conducted a systematic review of the psychometric evidence related to the CMS. Regarding the content validity of the score studies, they suggested that the description in the original publication was insufficient to accomplish standardization between centers and evaluators. Recent publications of the translation and cultural adaptation of the CMS into Danish,\(^ {10} \) Turkish,\(^ {12} \) and
Greek\textsuperscript{14} provided a standardized test protocol incorporating the new and modified guidelines provided by Constant and Murley\textsuperscript{9} together with the reliability and validity assessments. Even though the CMS is widely used in Italy, there has been no translation or cultural adaptation until now. The aim of this study was to validate the CMS after a process of Italian translation and cultural adaptation. This process was carried out in accordance with the international guidelines,\textsuperscript{28} and it was followed by establishing the internal consistency, reliability, and construct validity according to the international literature recommendations.\textsuperscript{30} Some modifications were implemented in the translation to increase the specificity of the CMS questionnaire for detecting shoulder pain in the Italian population. The preliminary CMS-IT was extensively analyzed by physicians, orthopedic surgeons, physical therapists, and an occupational therapist.

To validate the scale, a 7-month study was carried out at two Italian clinics. Statistical tests were performed to assess its reliability and validity, and the results were positive. The internal consistency showed a Cronbach’s α value of 0.81, which resembles those of previous studies: 0.74 in a Chinese study\textsuperscript{5} and 0.85 in a Brazilian Portuguese study.\textsuperscript{11} The means of the ICCs with regard to the inter-rater and intra-rater reliabilities were 0.994 (range = 0.987–0.997, \(p < 0.001\)) and 0.963 (range = 0.926–0.982, \(p < 0.001\)), respectively. Based on these results, it is possible to say that the reliability of the scale was very good and that the test had good reproducibility. In fact, performing the tests at a later time with no significant changes in the patient’s health status produced results consistent with the ones obtained from the first administration. These are comparable to the previously reported results from the Turkish (0.87)\textsuperscript{12} and Danish (0.94)\textsuperscript{10} versions. The inter-rater reliability was also very good, showing that the test was not rater-dependent. In addition, this study confirmed the correlation between the CMS-IT and VAS (\(-0.55\)).\textsuperscript{26} This negative correlation corresponded to the characteristics of each scale: a high VAS score indicates a more serious clinical picture, whereas a high CMS score corresponds to better shoulder function quality. With regard to the convergent validity, the CMS-IT demonstrated a strong negative correlation with the Italian version of the DASH (\(r = -0.47, p \leq 0.001\))\textsuperscript{2} and the previously reported Brazilian Portuguese validation study (\(r = -0.82, p \leq 0.001\)).\textsuperscript{12} The Danish version\textsuperscript{10} of the CMS also demonstrated a strong correlation with the Oxford Shoulder Score (\(r = 0.76\)). In spite of the similar values found in both adapted versions, our reference standard score was the Italian DASH, which is widely used in myriad shoulder conditions, while the Oxford Shoulder Score is more suited to assess surgical populations and proximal humerus fractures.

**Study limitations**

This study did have certain limitations. For example, the responsiveness, minimum detectable change, and minimally important change were not determined. Moreover, a sensitivity analysis of the CMS-IT was not carried out, which may have involved more factors, such as the effectiveness of the intervention as well as the type and severity of the disease. Future studies should be conducted to analyze and confirm these properties.

**Conclusions**

Despite the fact that the CMS has received much criticism for several issues, it is still widely used to assess the functional status of patients suffering from shoulder disorders. This study revealed that the CMS-IT contains good internal consistency and good test–retest reliability. Moreover, the results suggested that the CMS-IT questionnaire is a reliable and valid tool for assessing shoulder problems in the Italian population, and it deserves broad applications in both clinical practice and research contexts. This scale can also be used as an alternative to the current “gold standard instrument” VAS and DASH.

**Authors’ Note**

Donatella Valente is also affiliated with IRCSS Neuromed, Via Atinense, Pozzilli, IS, Italy.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Ethical approval**

The authors certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during the course of this research. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008.

**Informed consent**

Informed consent was obtained from all participants for being included in the study.

**Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iD**

Giovanni Galeoto  
https://orcid.org/0000-0002-9043-5686

**References**

1. Noten S, Struyf F, Lluch E, et al. Central pain processing in patients with shoulder pain: a review of the literature. *Pain Pract* 2017; 17(2): 267–280.
2. Slobogean GP and Slobogean BL. Measuring shoulder injury function: common scales and checklists. *Injury* 2011; 42(3): 248–252.
3. Fayad F, Mace Y and Lefevre-Colau MM. Shoulder disability questionnaires: a systematic review. Ann Readapt Med Phys 2005; 48(6): 298–306.

4. Constant CR and Murley A. A clinical method of functional assessment of the shoulder. Clin Orthop Relat Res 1987; 214: 160–164.

5. Angst F, Schwzyer HK, Aeschlimann A, et al. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version (QuickDASH), shoulder pain and disability index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, Constant (Murley) score (CS), simple shoulder test (SST), Oxford shoulder score (OSS), shoulder disability questionnaire (SDQ), and Western Ontario shoulder instability index (WOSI). Arthritis Care Res 2011; 63(Suppl 11): S174–S188.

6. Kirkley A, Griffin S and Dainty K. Scoring systems for the functional assessment of the shoulder. Arthroscopy 2003; 19(10): 1109–1120.

7. Barra-López ME. El test de Constant–Murley. Una revisión de sus características. Rehabilitation 2007; 41(5): 228–235.

8. Constant CR, Gerber C, Emery RJH, et al. A review of the constant score: modifications and guidelines for its use. J Shoulder Elbow Surg 2008; 17(2): 355–361.

9. Vrotsou K, Ávila M, Machón M, et al. Constant–Murley score: systematic review and standardized evaluation in different shoulder pathologies. Qual Life Res 2018; 27(9): 2217–2226.

10. Moeller AD, Thorsen RR, Torabi TP, et al. The Danish version of the modified Constant–Murley shoulder score: reliability, agreement, and construct validity. J Orthop Sport Phys Ther 2014; 44(5): 336–340.

11. Barreto RPG, Barbosa MLL, Balbinotti MAA, et al. Versão brasileira do Constant–Murley Score (CMS-BR): validade convergente e de construto, consistência interna e unidimensionalidade. Rev Bras Ortop 2016; 51(5): 515–520.

12. Çelik D. Turkish version of the modified Constant–Murley shoulder score and standardized test protocol: reliability and validity. Acta Orthop Traumatol Turc 2016; 50(1): 69–75.

13. Yao M, Yang L, Cao Z-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(1): 178.

14. Ntourantonis D, Panagopoulos A, Iliopoulos I, et al. Translation, cultural adaptation, validity and reliability of the Greek version of the modified Constant Score. JSES Open Access 2017; 1(1): 45–50.

15. Murena L, Vulcano E, D’Angelo F, et al. Italian cross-cultural adaptation and validation of the Oxford shoulder score. J Shoulder Elbow Surg 2010; 19(3): 335–341.

16. Massai P, Colalelli F, Sansoni J, et al. Reliability and validity of the geriatric depression scale in Italian subjects with Parkinson’s disease. Parkinsonism Dis 2018; 2018: 7347859.

17. Covotta A, Gagliardi M, Berardi A, et al. Physical activity scale for the elderly: translation, cultural adaptation, and validation of the Italian version. Curr Gerontiol Geriatr Res 2018; 2018: 8294568.

18. Berardi A, Biondillo A, Márquez MA, et al. Validation of the short version of the Van Lieshout Test in an Italian population with cervical spinal cord injuries: a cross-sectional study. Spinal Cord 2018; 57: 339–345.

19. Dattoli S, Colucci M, Soave MG, et al. Evaluation of pelvis postural systems in spinal cord injury patients: outcome research. J Spinal Cord Med 2018; 43(2): 185–192.

20. Galeoto G, Colucci M, Guarino D, et al. Exploring validity, reliability, and factor analysis of the Quebec user evaluation of satisfaction with assistive technology in an Italian population: a cross-sectional study. Occup Ther Heal Care 2018; 32(4): 380–392.

21. Berardi A, Dhrami L, Tofani M, et al. Cross-cultural adaptation and validation in the Italian population of the wolf motor function test in patients with stroke. Funct Neurol 2018; 33: 229–253.

22. Savona A, Ferralis L, Saffioti M, et al. Evaluation of intra-and inter-rater reliability and concurrent validity of the Italian version of the Jebson–Taylor Hand Function Test in adults with rheumatoid arthritis. Hand Ther 2019; 24: 48–54.

23. Galeoto G, Iori F, Santis RD, et al. The outcome measures for loss of functionality in the activities of daily living of adults after stroke: a systematic review. Top Stroke Rehabil 2019; 26(3): 236–245.

24. Ruggeri M, Lasalvia A, Nicolaou S, et al. The Italian version of the Cumberwell Assessment of Need (CAN), an interview for the identification of needs of care. Epidemiol Psychiatr Soc 1999; 8: 135–139.

25. Castiglia SF, Galeoto G, Lauta A, et al. The culturally adapted Italian version of the Barthel Index (IcaB1): assessment of structural validity, inter-rater reliability and responsiveness to clinically relevant improvements in patients admitted to inpatient rehabilitation centers. Funct Neurol 2017; 22: 221–228.

26. Bijur PE, Silver W and Gallagher EJ. Reliability of the visual analog scale for measurement of acute pain. Acad Emerg Med 2001; 8(12): 1153–1157.

27. Padua R, Padua L, Ceccarelli E, et al. Italian version of the Disability of the Arm, Shoulder and Hand (DASH) questionnaire. Cross-cultural adaptation and validation. J Hand Surg Am 2003; 28(2): 179–186.

28. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Heal 2005; 8(2): 94–104.

29. Conboy VB, Morris RW, Kiss J, et al. An evaluation of the Constant–Murley shoulder assessment. J Bone Joint Surg Br 1996; 78(2): 229–232.

30. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010; 19(4): 539–549.

31. Roy JS, Macdermid JC and Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. Arthritis Care Res 2009; 61(5): 623–632.