Musculoskeletal Science and Practice 43 (2019) 37–44

Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

journal homepage: www.elsevier.com/locate/msksp

Original article

# Cross-cultural adaptation and validation of the Argentine "American Shoulder and elbow surgeons, patient self-report section" questionnaire

Pablo Oscar Policastro[a,b,*], Andrés Pierobon[a], Joaquín Pérez[a], Gabriel Adrián Novoa[a], Melina Calvo Delfino[a], Malka Eugenia Sajfar[a], Sandra Salzberg[a], Candela Carmody[a], Javier Hernán Dorado[a], Ignacio Raguzzi[a], Santiago Soliño[a], Eliana Cecilia Pérez Calvo[a]

[a] Physical Therapy Unit, Durand Hospital. Buenos Aires City, Argentina
[b] KINÉ- Kinesiología Deportiva y Funcional Sports Clinic, Buenos Aires City, Argentina

ARTICLE INFO

*Keywords:*
Outcome measures
Shoulder
Validation studies

ABSTRACT

*Background:* American Shoulder and Elbow Surgeons questionnaire (ASES-p) has been translated into Spanish, but it has not been adapted to the Argentine population yet. Although Spain and Argentina speak the same language, linguistic differences between Spanish-speaking countries may affect the interpretation of the different items included in the questionnaire.
*Objective:* To conduct the translation, cross-cultural adaptation and validation of the self-report section of the ASES-p into Argentine Spanish for patients with musculoskeletal shoulder disorders, and to assess its psychometric properties.
*Design:* Study of diagnostic accuracy/assessment scale.
*Method:* The study was carried out in three consecutive phases: translation, cross-cultural adaptation and validation for its use in Argentina. In the third phase, we used the ASES-p, Short Form 36 (SF-36), EuroQol-5D (EQ-5D), and Disabilities of the Arm, Shoulder and Hand (DASH) questionnaires, and the Global Rating of Change (GROC) scale.
*Results:* One hundred three participants completed a set of questionnaires on two occasions and were included in the final analysis. The time taken to answer and score the questionnaire was 118 and 52 s, respectively. Neither a ceiling nor a floor effect was observed. Cronbach's alpha coefficient was 0.85. Intraclass correlation coefficient was 0.83. A significant correlation was found between the DASH, the GROC and various SF-36 subscales. There were strong indices of concurrent-cross validation, longitudinal validity, and construct validity. The ASES-p questionnaire showed a minimal clinically important difference (MCID) value of 7.88 points.
*Conclusion:* Some psychometric properties in reliability and validity were acceptable in the Argentine version of the ASES-p questionnaire.

## 1. Introduction

Shoulder musculoskeletal disorders are common medical conditions which cause pain, mobility problems and disability. These conditions may limit work activity and may affect quality of life, causing a significant socio-economic burden (Luime et al., 2004; Schmidt et al., 2014). In recent years, care models have given patients a primary role in decision making; hence, self-reported outcome measures are becoming increasingly recognized and used in medical research. A systematic review compared the psychometric properties of different self-reported questionnaires to assess shoulder musculoskeletal disorders. The American Shoulder and Elbow Surgeons questionnaire patient section (ASES-p) obtained the highest general score and was one of the best in all the evaluated items (Schmidt et al., 2014).

The ASES-p has two domains: pain and function (activities of daily living, ADLs) (Richards et al., 1994). Both domains were validated initially in 2002 (Michener et al., 2002). After that, the scale has been validated in different countries of North and South America, Europe and Africa, demonstrating good psychometric properties (Michener et al., 2002; Goldhahn et al., 2008; Padua et al., 2010; Yahia et al.,

* Corresponding author. 1785 Castillo, Buenos Aires City, 1414, Argentina.
*E-mail address:* pablopolicastro@gmail.com (P.O. Policastro).
@policastrito (P.O. Policastro), @andres_pierobon (A. Pierobon), @joaquin_perez9 (J. Pérez), @gabriel_n89 (G.A. Novoa), @calvomelina (M. Calvo Delfino), @JavierhDorado (J.H. Dorado), @pikeraguzzi (I. Raguzzi), @sskinesio (S. Soliño)

https://doi.org/10.1016/j.msksp.2019.05.010
Received 20 February 2019; Received in revised form 27 May 2019; Accepted 29 May 2019

2011; Moser et al., 2012; Celik et al., 2013; Piitulainen et al., 2014). Also the possibility to include two activities reported by the patient give the opportunity of more personalized assessment.

The questionnaire has been translated into Spanish (Vrotsou et al., 2016), but it has not been adapted to the Argentine population yet. Although Spain and Argentina speak the same language, linguistic differences between Spanish-speaking countries may affect the interpretation of the different items included in the questionnaire. This might mainly occur because some of the terms used in the Spanish version are not used in Argentina. For example, the equivalent for the English term "bra", translated as *sujetador* in Spain, is *corpiño* in Argentina. Differences in terminology could affect interpretation and therefore answers. The language within a country is an indirect indicator of culture. Ignoring specific language differences between countries could negatively influence the validation of the tool, since the differences in scores could be due to real differences in health status or to an incorrect interpretation (Ware et al., 1995). Additionally, it is important to have validated scales in our country in order to use them in future research studies and in daily clinical practice as an outcome measure. Therefore, the aim of this study was to conduct the translation, cross-cultural adaptation and validation of the ASES-p questionnaire into Argentine Spanish for patients with shoulder musculoskeletal disorders.

## 2. Materials and methods

Before starting with the study, the protocol was officially authorized by the American Shoulder and Elbow Surgeons Society, the original developer. This study was granted by the Research Ethics Committee of Durand Hospital, Autonomous City of Buenos Aires, Argentina, and it consisted of three consecutive phases: translation, cross-cultural adaptation and validation. Written consent was obtained from each participant before data collection.

### 2.1. Translation, cross-cultural adaptation and pilot study

These phases were performed following a previously described guideline (Beaton et al., 1976). Initially, the questionnaire was independently translated into Spanish by two translators who were native Argentine Spanish speakers, and only one of them had knowledge about this field. Then, the translators and the authors of the study reached a consensus for the first Argentine Spanish version of the questionnaire. After that, this version was independently back-translated into English by two bilingual translators who were native English speakers living in Argentina. These translators were unaware of the original questionnaire. A preliminary version of the questionnaire was accepted by a Committee constituted by twelve physical therapists and specialists in shoulder disorders, who were the authors of the study; two translators; a psychologist; a sociologist; and a specialist in statistics. The data as name, sex, age, dominant shoulder, day and time of evaluation, score of the scales, time in scoring and completing ASES-p, were collected on a specific written form designed for this study and stored on a database. In the pilot test, a survey was given to 30 study participants who met the inclusion and exclusion criteria listed below. This test aimed at identifying any comprehension difficulties or any other obstacles to complete it (Terwee et al., 2007).

### 2.2. Validation

Participants were recruited consecutively and prospectively in the outpatient physical therapy unit of the Durand Hospital. The inclusion criteria were: being more than 18, being Argentinian, having medical referral diagnosis of shoulder musculoskeletal disorder, and having the appropriate informed consent signed. Certain participants were excluded from the study: patients with functional impairment or pain related to systemic diseases (e.g. rheumatoid arthritis and collagen disorders), patients with neoplasms or infection of another structure of the upper limb, patients with neurological disorders, patients with communication or understanding difficulties, illiterate patients, patients who refused to take part in the study, and patients with bilateral shoulder pain. This last is because the measures of responsiveness the symptoms of one shoulder could affect the perception of the symptoms of the other shoulder as well the performance of some functional activities. Participants who refused to continue the study, abandoned treatment or completed only the initial evaluation (T1) were eliminated.

### 2.3. ASES-p questionnaire

The ASES-p scale is composed of 11 items, divided in 2 subscales: pain (1 item) and function (10 items). Pain intensity is measured on a 100 mm Visual Analogue Scale (VAS). The function subscale evaluates the ability to perform ten Activities of Daily Living (ADLs) on a Likert scale (0 unable to do, 1 very difficult to do, 2 somewhat difficult, 3 not difficult (Michener et al., 2002). In order to obtain the results, the proposed equation in the original questionnaire was used (Richards et al., 1994).

### 2.4. Other scales

The Short Form 36 (SF-36) was used to measure health status, and the EuroQol-5D (EQ-5D) was used to measure quality of life. Both scales have been validated for the Argentine population (Augustovski et al., 2008, 2009).

Furthermore, the Spanish version of the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire was used to assess the physical function of the upper limb. The Spanish version was also adapted to the Argentine population (Ellis et al., 2007).

Finally, a 15 point-global rating of change (GROC) scale was used to determine clinically important changes (Kamper et al., 2009).

### 2.5. Procedure

During the initial evaluation (T1), the participants completed the following questionnaires: the ASES-p, SF-36, EQ-5D and DASH. Within 24–72 h of completing T1, the participants completed the second evaluation (T2), consisting of the GROC scale and ASES-p questionnaire. During the period of time between T1 and T2, the participants did not receive treatment in order to maintain clinical stability. The third evaluation (T3) was carried out at the patient's discharge or four weeks after T1, whichever was first. During T3, the participants completed the ASES-p, DASH, EQ-5D and SF-36 questionnaires, and the GROC scale.

The order of administration of the questionnaires was randomized as in previous studies (Odetunde et al., 2018). Sealed and opaque envelopes were used to maintain concealment. Researchers were trained in the administration and scoring of the questionnaires before the study began.

### 2.6. Sample size

Considering a 20% loss to follow-up, at least 125 participants were recruited. The data collection was maintained until completing data of 100 participants. This is considered to be an excellent sample size to evaluate psychometric properties (Hobart et al., 2012; Terwee et al., 2012).

Data collection was classified into "complete", "incomplete" or "eliminated participants". Participants who did not complete T2 were eliminated. Participants who completed evaluation T1 and T2 but who did not complete T3 were considered "incomplete". Both incomplete and complete data were analyzed for the purposes of assessing the existence of any differences between both groups.

### 2.7. Statistical analysis

Continuous variables with normal distribution are presented as means and standard deviation (SD). Otherwise, the median and the interquartile range (IQR) were used. Categorical variables were expressed in numbers and percentages. The Shapiro-Wilk test was used to determine the distribution of the sample. A p value ≤ 0.05 was considered statistically significant. The software IBM SPSS Macintosh, version: 22.0 (IBM Corp, Armonk, NY, USA) was used for data analysis.

### 2.8. Content validity

When more than 15% of participants achieved the lowest or highest possible score of the questionnaire, ceiling and floor effects were considered present (Terwee et al., 2007).

### 2.9. Reliability

#### 2.9.1. Internal consistency

Internal consistency (IC) was assessed by the Cronbach's alpha coefficient (Cronbach's α) and it was considered acceptable when the coefficient ranged from 0.7 to 0.95 (Terwee et al., 2007). Item-item and item-total correlations were assessed by the Spearman correlation coefficient. Indices between 0.1 and 0.8 were acceptable for item-item correlations, and p value ≤ 0.05 was acceptable for item-total correlations (Clark and Watson, 1995; Tavakol et al., 2011; McHugh ML, 2012; Jean-Baptist du Prel, 2009).

#### 2.9.2. Test – retest

Test-retest reliability was determined by calculating the intraclass correlation coefficient (ICC) with confidence intervals of 95%. Participants who indicated no changes (GROC -2 to +2) in T2 were analyzed (Mokkink et al., 2010; Koo and Li, 2016).

#### 2.9.3. Validity

Pearson and Spearman (rho) correlation coefficients were used, as applicable. Correlation coefficients > 0.70, between 0.4 and 0.70, and < 0.40 were regarded as strong, moderate or weak, respectively (Post MW, 2016). To evaluate concurrent-cross validation, different hypotheses were formulated among assessment instruments, and previously reported criteria were followed (De Boer et al., 2004) (APPENDIX 1). Longitudinal validity was established by correlating changes in T1 and T3 ASES-p scores with changes in the DASH, EQ-5D, and SF-36 questionnaires (Husted et al., 2000).

Construct validity was determined by comparing the variables (means) of changes in ASES-p scores between T1 and T3 in the group of participants who improved (GROC +3 to +7), between T1 and T2 in the group of participants who remained unchanged (GROC -2 to +2), and by comparing the changes in scores of both groups. Pre and post-test values were compared by using the dependent *t*-test or Wilcoxon test, as applicable. Also, the independent-samples *t*-test or the Mann–Whitney *U* test, as applicable, were used to determine the differences between the groups.

In addition, scores of the ASES-p questionnaire in T3 were compared in the following groups: Group without changes (GROC -2 to +2), group with minimal changes (GROC +3 and + 4) and group with significant changes (GROC +5 to +7). The differences in mean values in the ASES-p of the three groups were determined by analyzing the covariance (ANCOVA). ASES-p scores in T1 were used as a covariate.

#### 2.9.4. Responsiveness

The analysis was made by using an external anchor (GROC) to determine responsiveness, in accordance with methods published elsewhere (Crosby et al., 2003; De Vet et al., 2006). Sensitivity to change was determined by calculating the Minimal Detectable Change (MDC) with the formula: 1.96 x √2 x SEM (the standard error of measurement) (De Vet et al., 2006). The standardized response mean (SRM) was calculated to determine the magnitude of change (Guyatt et al., 1987).

The Receiver Operating Characteristic (ROC) curve was used to assess the effectiveness of the ASES-p questionnaire, in accordance with Youden's index for rating diagnostic tests, to discriminate between those participants who improved and those who did not (Youden, 1950). ROC curves were plotted with pROC package, version 1.10.0 (R Core Team, 2017).

#### 2.9.5. Minimal clinically important difference (MCID) and substantial clinical benefit (SCB)

The MCID was calculated by using an external anchor-based approach. This approach compares the changes in ASES-p scores between T1 and T3 for participants who indicated minimal changes (GROC +3 and + 4), and between T1 and T2 for participants who indicated no changes (GROC -2 to +2). The SCB based on the anchor approach was calculated in a similar way to the MCID; however, this approach compares the change in ASES-p scores between T1 and T3 for participants who indicated a meaningful improvement (GROC ≥ +5) (Kon et al., 2014; Werner et al., 2016).

## 3. Results

### 3.1. Translation, cross-cultural adaptation and pilot study

The members of the Committee found no discrepancies during the translation and back-translation process of the ASES-p. During the pilot test, no participant reported difficulties in understanding the questionnaire. Therefore, since there were no modifications made, the final version of the ASES-p questionnaire was developed, and the results arising from the pilot test were included in the final analysis (APPENDIX 2).

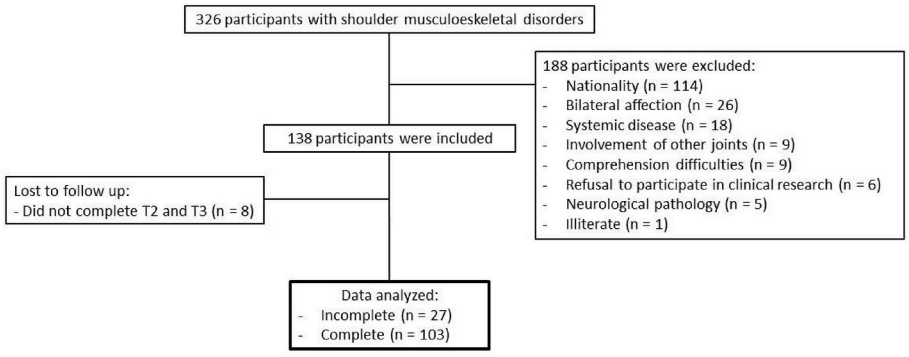


**Fig. 1.** Flow chart.

**Table 1**
Demographic characteristics of the sample.

| Variable | Analyzed n = 130 | Eliminated n = 8 |
|---|---|---|
| Female, n (%) | 73 (56.24) | 6 (75) |
| Age in years, median (IQR) | 55 (37.51–61) | 52.70 (10) |
| Medical Diagnosis, n (%) | | |
| Shoulder pain | 27 (20.82) | 4 (50) |
| Fracture | 11 (8.54) | 3 (37.50) |
| Tendinopathy | 56 (43.14) | 0 (0) |
| Periarthritis | 5 (3.87) | 0 (0) |
| Luxation | 11 (8.55) | 0 (0) |
| Arthrosis | 1 (0.88) | 1 (12.50) |
| Surgery | 3 (2.31) | 0 (0) |
| Other | 16 (12.33) | 0 (0) |
| Pain duration [a], weeks | 8 (4–16) | 8.5 (4–25) |
| Employed, n (%) | 85 (65.49) | 5 (62.50) |
| Right handed, n (%) | 121 (93.11) | 8 (100) |
| Affected Shoulder, right, n (%) | 72 (55.41) | 6 (75) |
| Educational level, n (%) | | |
| Primary Incomplete | 4 (3.16) | 0 (0) |
| Primary Complete | 23 (17.75) | 2 (25) |
| Secondary Incomplete | 21 (16.22) | 3 (37.50) |
| Secondary Complete | 37 (28.59) | 2 (25) |
| Tertiary/University | 45 (34.62) | 1 (12.50) |
| Self-report questionnaires, | | |
| ASES-p | 43.50 (33.10–60.30) [a] | 31.3 (17.40) [b] |
| DASH | 45.7 (30–62) [a] | 51.8 (27.80) [b] |
| SF-36 | 57.2 (16.90) [b] | 50.21 (32.30–78.50) [a] |
| EQ-5D | 72 (52–88) [a] | 57 (35–73.70) [a] |

**References: ASES-p:** American Shoulder and Elbow Surgeons questionnaire patient section; **DASH:** Disabilities of the Arm, Shoulder and Hand; **EQ-5D**: EuroQol-5D; **n:** number**; SF-3**6: Short Form 36**; %:** percentage.
[a] median and interquartile range.
[b] mean and standard deviation.

### 3.2. Validation

The flow chart of the validation process is shown in Fig. 1.

The demographic characteristics of the analyzed participants are shown in Table 1.

The median (IQR) of the time taken to complete the ASES questionnaire was 118 (81.54–170.71) seconds. The time taken by the evaluators to score the questionnaire was 52 (44–63) seconds.

A total of 138 participants answered the questionnaire in T1, the mean (SD) obtained was 45.50 (19.10) points. None of the participants reported extreme scores, that is, no floor or ceiling effects were observed. Minimum and maximum scores reported were 6 and 97.10, respectively, Cronbach's alpha coefficient was 0.85 when all items were included. After the exclusion of item 11 in the scale, Cronbach's alpha coefficient was 0.90. The correlation coefficients between the sub-scales function and pain were weak (rho = 0.28). Item-item correlations were between 0.28 and 0.67. The minimum value was for item 11, and the maximum value was for item 7. The correlations related to items 9, 10 and 11 were below the values suggested by previous studies (Table 2).

ICC was 0.83 (95% confidence interval (95% CI): 0.77, 0.88.) Values between 0.77 and 0.88 indicate good reliability.

### 3.3. Concurrent cross validation

Most hypotheses (84.60%) were confirmed, and a strong concurrent-cross validation was found (Table 3).

### 3.4. Longitudinal validity

Half of the hypotheses (50.00%) were confirmed, and a moderate longitudinal validity was found (Table 3).

**Table 2**
Internal consistency.

| Subscale – Function | Correlation – Item - Total |
|---|---|
| Item 1: Put on a coat | 0.55 |
| Item 2: Sleep on your painful or affected side | 0.62 |
| Item 3: Wash back/do up bra in back | 0.56 |
| Item 4: Manage toileting | 0.50 |
| Item 5: Comb hair | 0.46 |
| Item 6: Reach a high shelf | 0.65 |
| Item 7: Lift 10 pounds above shoulder | 0.67 |
| Item 8: Throw a ball overhand | 0.54 |
| Item 9: Do usual work- List: | 0.42 |
| Item 10: Do usual sport/hobby- List: | 0.38 |
| **Subscale – Pain intensity** | |
| Item 11: How bad is your pain today (place a line)? | 0.28 |
| Cronbach's Alpha – Total | 0.85 |
| Cronbach's Alpha – Subscale | 0.90 |

Item-total correlations were estimated with Spearman's correlation coefficient.

**Table 3**
Concurrent-cross and longitudinal validity.

| Questionnaires | ASES-p Correlation Concurrent Validity | ASES-p Correlation Longitudinal Validity |
|---|---|---|
| DASH | −0,75[a] | −0.61[a] |
| SF-36 Role Physical | 0.47 | 0.27 |
| SF-36 Physical Functioning | 0.6 | 0.48 |
| SF-36 Bodily Pain | 0,58 | 0.37 |
| SF-36 Physical Component Summary PCS | 0.65[a] | 0.47[a] |
| EQ-5D Self-Care | −0.42 | −0.34 |
| EQ-5D Usual Activities | −0.46 | −0.37 |
| EQ-5D Pain/Discomfort | −0.39 | −0.21 |
| SF-36 Role Emotional | 0.39[a] | 0.11 |
| SF-36 Mental Health | 0.24 | 0.29 |
| SF-36 Mental Component Summary MCS | 0.41 | 0.36[a] |
| EQ-5D Mobility | −0.29 | −0.005 |
| EQ-5D Anxiety/Depression | −0.24 | −0.32 |
| GROC | | 0.52 |

[a] Pearson's correlation coefficient. If not, Spearman's correlation coefficient.

### 3.5. Construct validity

The paired sample *t*-test showed that changes in ASES-p scores were not significantly different between T1 and T2 (unchanged group $p = 0.66$). However, there is a significant difference between T1 and T3 in the group with significant changes ($p < 0.001$). The Mann-Whitney *U* Test applied to independent samples showed statistically significant differences between the no-change group and the significant-change group ($p < 0.001$).

In the covariance analysis of the three groups (no-change, minimum-change, and significant-change), statistically significant differences were found in ASES-p T3 scores ($p < 0.001$) (Table 4).

After the post-hoc analysis (Bonferroni correction), statistically significant differences were observed in all group comparisons, demonstrating evidence in support of construct validity ($p < 0.001$).

**Table 4**
ASES-p subgroups.

| | n | ASES-p T1 mean (SD) | ASES-p T3 mean (SD) |
|---|---|---|---|
| No- change group | 10 | 51.20 (26.11) | 45.59 (23.87) |
| Minimum- change group | 32 | 42.41 (14.75) | 61.55 (13.63) |
| Significant- changed group | 58 | 47.22 (21.43) | 80.03 (16.09) |

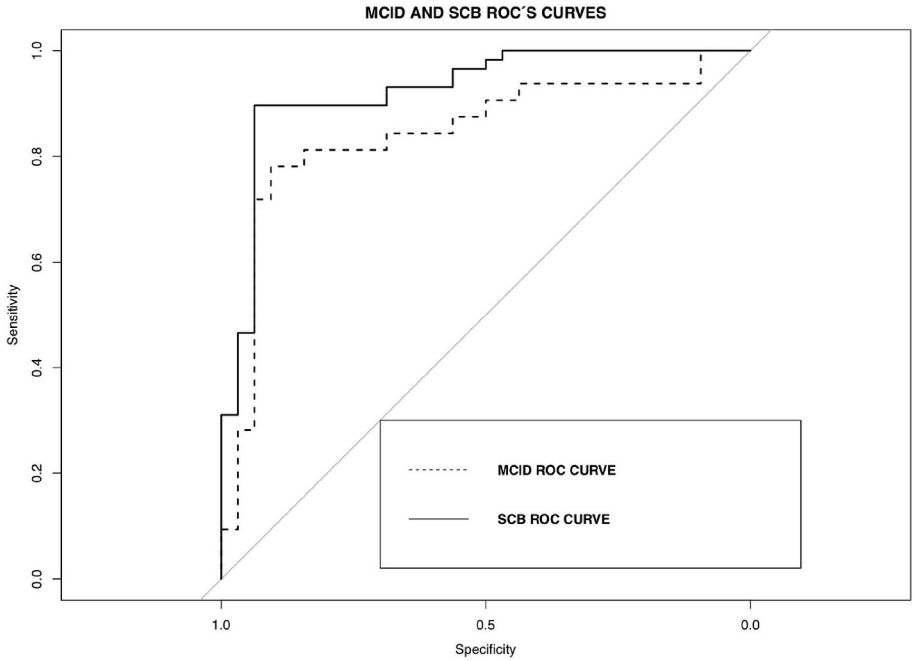


**Fig. 2.** Receiver operating curves. Minimal clinically important diference (MCID) and substantial clinical benefit (SCB).

### *3.6. Sensitivity to change*

SEM and $MDC_{95}$ values were 7.56 and 17.60, respectively. Fig. 2 shows the area under the ROC curve concerning MCID and SCB, respectively. The MCID was 7.88, and the SCB was 9.24 with a standardized response mean (SRM) > 0.8 (cutoff point).

## 4. Discussion

In the present study, the translation, cross-cultural adaptation and validation of the self-report section of the American Shoulder and Elbow Surgeons questionnaire (ASES-p) into Argentine Spanish were conducted. This instrument is used worldwide and has demonstrated good psychometric properties (Michener et al., 2002; Goldhahn et al., 2008; Padua et al., 2010; Yahia et al., 2011; Moser et al., 2012; Celik et al., 2013; Piitulainen et al., 2014).

The Argentine version of the ASES-p questionnaire has shown to be a viable instrument for clinical assessment, requiring less than 2 min to complete it and only 1 min to score the questionnaire. This is the first study to report the preceding information; consequently, the comparison with other samples is not possible. Furthermore, as in previous validations, neither a ceiling nor a floor effect was observed (Kocher et al., 2005; Padua et al., 2010; Celik et al., 2013; Piitulainen et al., 2014; Vrotsou et al., 2016).

The internal consistency was considered acceptable, which is similar to previous validations (Michener et al., 2002; Padua et al., 2010; Celik et al., 2013; Piitulainen et al., 2014). Regarding item-item and item-total correlations, items 9, 10 and 11 were below the acceptable values. This may have occurred because the instructions to complete those items required a greater degree of interpretation and understanding. For items 9 and 10, a wide variety of responses was reported, as the participants had to answer based on a job and a sport of their choice. In addition, the activities chosen may imply different physical demands for the upper limb, which may require a minimum range of motion (e.g. watching tv) or a maximum range of motion (e.g. playing tennis). Item 11 refers to the assessment of pain by a Visual Analogue Scale (VAS). The low correlation value obtained between this question and the total score may be related to the different factors affecting self-reported pain, such as previous experience of pain, age, emotional factors, context and what each patient understands as pain. Each participant's line represents different pain intensity; that is, one participant may indicate severe pain, whereas a second participant may indicate moderate pain (Williamson and Hoggart, 2005; Reed et al., 2014). Future investigations could evaluate a shorter new version of ASES-p score avoiding some items.

Test-retest reliability was considered good (Koo and Li, 2016), which is in line with previous research findings (Michener et al., 2002; Piitulainen et al., 2014). It should be emphasized that previous studies do not agree with the required duration of time intervals to assess reliability. Even though in the Argentine version the time interval was 24–72 h, as previously reported (Michener et al., 2002; Yahia et al., 2011), there are other validations with greater time intervals (Kocher et al., 2005; Goldhahn et al., 2008; Padua et al., 2010; Moser et al., 2012; Celik et al., 2013; Piitulainen et al., 2014).

In order to determine the cross-validation of the questionnaire, the correlation with the DASH and SF-36 was analyzed. The ASES-p questionnaire was strongly correlated with the DASH (Goldhahn et al., 2008; Padua et al., 2010; Yahia et al., 2011). This could arise because the DASH assesses symptoms and the physical function of the upper limb, and activities that may be related to similar activities in the ASES-p. In accordance with the hypothesis, a moderate correlation was found with the physical role domain of the SF-36 questionnaire.

The longitudinal validity of the ASES-p was determined by assessing the correlation with DASH and GROC, which was moderate in both cases. Thus, the scale was found to be a useful tool to detect post-treatment changes.

MCID is defined as the minimal score or difference that may be perceived by patients as clinically beneficial or harmful. Although there is no agreement concerning the best way to calculate the MCID, the use of external anchors is considered a valid alternative, and they are the most commonly used by scholars (Revicki et al., 2008; Minagawa et al., 2013; Tashjian et al., 2017). In this study, in order to calculate the MCID, patients who scored GROC +3 and + 4 were included. The area under the ROC curve allows to measure the threshold value of sensitivity and specificity (Tashjian et al., 2010). Only one of the ASES-p validations obtained the MCID (6.40 points). The curve reported 75% of sensitivity and 91% of specificity. The area under the ROC curve was 0.81 (Michener et al., 2002). In the Argentine version, the MCID was

7.88, and sensitivity and specificity values as well as the area under the ROC curve were very similar to those previously reported. Nevertheless, Michener et al. analyzed participants with minimal clinically important changes as well as those with substantial benefits; this differs from our study (Michener et al., 2002). Other authors found greater clinical changes, which may be attributed to the use of different cut off points to determine the substantial clinical benefit, or to the analysis of a sample size exclusively surgical in nature (Tashjian et al., 2010, 2017). A remarkable finding in our study was that the MDC was greater than the MCID. Even though this is unusual, it has been reported elsewhere (Mintken et al., 2009). The MCID can be less or greater than the MDC given that the first one is based on a patient response-anchored method (MCID), while the second one is calculated as a statistical threshold (MDC) (Wright et al., 2012). A value above the MCID but below the MDC should be interpreted as a clinically important change, which cannot be distinguished from measurement error.

To our knowledge, this is the first study to report the SCB of the ASES-p questionnaire in a population with shoulder pain of different etiologies. In a previous study the SCB ranged from 12.00 to 36.60 (Werner et al., 2016). However, that analysis was a retrospective review and the study participants included were patients undergoing shoulder arthroplasty. This limits generalization of the results to populations with other medical conditions (Werner et al., 2016).

It is highlighted that, in the first place, the implementation of a pilot test made it possible to verify that participants understood the questionnaire. Also, the evaluators were trained in the administration of questionnaires. Secondly, the sample size used in the study was satisfactory to evaluate the psychometric properties of an assessment instrument. Lastly, both the inclusion of a wide range of pathologies and the age range analyzed allow the sample size to be largely representative of the shoulder pain population. However, there were some limitations. Data collection was made in only one health center, thus, the characteristics of shoulder disorders in other diverse contexts might be different from those assessed in this study. Another limitation was the time interval to assess reliability that might have been too short. Two weeks is often considered appropriate for the evaluation of self-reported outcome measures (Prinsen et al., 2018). Finally, the lack of sample with bilateral shoulder pain would not allow to extrapolate the results to this population.

## 5. Conclusion

In the current study, we conducted the translation, cross-cultural adaptation and validation of the self-report section of the ASES-p questionnaire into Argentine Spanish. The current study allowed us to use the Argentine version of the ASES-p in the Argentine population and provided foundations to further investigate other psychometric properties to identify the optimal Patient Reported Outcome Measure for shoulder impairments.

## Ethical committee aproval

This study was approval by the Ethic's Committee of Durand Hospital, Buenos Aires City Argentina.

## Conflicts of interest

None.

## Statement of financial disclosure

These authors, their immediate family, and any research foundation with which they are affiliated did not receive any financial payments or other benefits from any commercial entity related to the subject of this article.

## Acknowledgements:

We would like to acknowledge to Adriel Chara, María Francisca Dominguez, Emiliano Navarro, Daniela Gilgado, Clarisa Schwerdt, María Eugenia Pereira, Agustina Ciarlantini, Mauro Andreu, David Logerstedt, Lori Michener, Physical Therapy Unit of Durand Hospital for their valuable contribution and recommendations.

## APPENDIX 1

Hypotheses formulated to determine concurrent-cross validation.

- There is a strong correlation between ASES-p scores and DASH scores.
- There is a moderate correlation between ASES-p scores and SF-36 scores concerning the following scales: role physical, physical functioning, bodily pain and physical component summary.
- There is a moderate correlation between ASES-p scores and EQ-5D scores concerning the following dimensions: self-care, usual activities and pain.
- There is a weak correlation between ASES-p scores and SF-36 scores concerning the following scales: role-emotional, mental health and mental component summary.
- There is a weak correlation between ASES-p scores and EQ-5D scores concerning the following dimensions: mobility and anxiety/depression.

Hypotheses formulated to determine longitudinal validity.

- There is a strong correlation between changes in ASES-p scores and the GRoC.
- There is a strong correlation between changes in ASES-p scores and DASH scores.
- There is a moderate correlation between changes in ASES-p scores and SF-36 scores concerning the following scales: role physical, physical functioning, bodily pain and physical component summary.
- There is a moderate correlation between ASES-p scores and EQ-5D scores concerning the following dimensions: self-care, usual activities and pain.
- There is a weak correlation between changes in ASES-p scores and SF-36 scores concerning the following scales: role-emotional, mental health and mental component summary.
- There is a low correlation between changes in ASES-p scores and EQ-5D scores of the following dimensions: mobility and anxiety/depression.

## APPENDIX 2

*Cuestionario ASES para evaluar la función*

En el siguiente cuadro, encierre con un círculo el número que indica su capacidad para realizar las siguientes actividades:
0 = Imposible 1 = Muy difícil 2 = Algo difícil 3 = Nada difícil.

| Actividad | Brazo derecho | Brazo izquierdo |
|---|---|---|
| 1. Ponerse un abrigo. | 0-1-2-3 | 0-1-2-3 |
| 2. Dormir sobre el lado dolorido o afectado. | 0-1-2-3 | 0-1-2-3 |
| 3. Lavarse la espalda/abrocharse el corpiño. | 0-1-2-3 | 0-1-2-3 |
| 4. Ir al baño e higienizarse | 0-1-2-3 | 0-1-2-3 |
| 5. Peinarse. | 0-1-2-3 | 0-1-2-3 |
| 6. Llegar an un estante alto. | 0-1-2-3 | 0-1-2-3 |
| 7. Levantar 4,5 kg por encima del hombro. | 0-1-2-3 | 0-1-2-3 |
| 8. Lanzar una pelota ubicando la mano por encima del hombro. | 0-1-2-3 | 0-1-2-3 |
| 9. Realizar las tareas habituales. Mencionar: | 0-1-2-3 | 0-1-2-3 |
| 10. Practicar los deportes/Hobbies habituales. Mencionar: | 0-1-2-3 | 0-1-2-3 |

¿Cuánto le duele hoy? (Marque sobre la siguiente línea)

0 Nada de dolor

10 Peor dolor imaginable

[(10 – puntaje EVA)] x 5 + [(5/3) x puntaje Función] = Total

## References


Augustovski, F.A., Lewin, G., Elorrio, E.G., Rubinstein, A., 2008 Dec. The Argentine-Spanish SF-36 Health Survey was successfully validated for local outcome research. J. Clin. Epidemiol. 61 (12), 1279–1284. https://doi.org/10.1016/j.jclinepi.2008.05.004.

Augustovski, F.A., Irazola, V.E., Velazquez, A.P., Gibbons, L., Craig, B.M., 2009 Jun. Argentine valuation of the EQ-5D health states. Value Health 12 (4), 587–596. https://doi.org/10.1111/j.1524-4733.2008.00468.x.

Beaton, D.E., Bombardier, C., Guillemin, F., Ferraz, M.B., 1976. Guidelines for the process of cross-cultural adaptation of selfreport measures. *Spine* (Phila Pa 25 (24), 3186–3191. 2000 Dec 15. https://doi.org/10.1097/00007632-200012150-00014.

Celik, D., Atalar, A.C., Demirhan, M., Dirican, A., 2013 Sep. Translation, cultural adaptation, validity and reliability of the Turkish ASES questionnaire. Knee Surg. Sport. Traumatol. Arthrosc. 21 (9), 2184–2189. https://doi.org/10.1007/s00167-012-2183-3.

Clark, L.A., Watson, D., 1995. Constructing validity: basic issues in objective scale development. Psychol. Assess. 7 (3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309.

Crosby, R.D., Kolotkin, R.L., Williams, G.R., 2003. Defining clinically meaningful change in health-related quality of life. J. Clin. Epidemiol. 56 (5), 395–407. https://doi.org/10.1016/S0895-4356(03)00044-1.

De Boer, M.R., Moll, A.C., de Vet, H.C., Terwee, C.B., Volker-Dieben, H.J., van Rens, G.H., 2004. Psychometric properties of vision-related quality of life questionnaires: a systematic review. Ophthalmic Physiol. Optic. 24, 257–273. https://doi.org/10.1111/j.1475-1313.2004.00187.x.

De Vet, H.C., Terwee, C.B., Knol, D.L., Bouter, L.M., 2006. When to use agreement versus reliability measures. J. Clin. Epidemiol. 59 (10), 1033–1039. https://doi.org/10.1016/j.jclinepi.2005.10.015.

Ellis, G., 2007. Review and Linguistic Validation of the Disabilities of the Arm, Shoulder and Hand Questionnaire (DASH), Spanish (Argentina). Oxford Outcomes, pp. 1–9.

Goldhahn, J., Angst, F., Drerup, S., Pap, G., Simmen, B.R., Mannion, A.F., 2008 Mar-Apr. Lessons learned during the cross-cultural adaptation of the American Shoulder and Elbow Surgeons shoulder form into German. J. Shoulder Elb. Surg. 17 (2), 248–254. https://doi.org/10.1016/j.jse.2007.06.027.

Guyatt, G., Walter, S., Norman, G., 1987. Measuring change over time: assessing the usefulness of evaluative instruments. J. Chronic Dis. 40 (2), 171–178. https://doi.org/10.1016/0021-9681(87)90069-5.

Hobart, J.C., Cano, S.J., Warner, T.T., Thompson, A.J., 2012. What sample sizes for reliability and validity studies in neurology? J. Neurol. 259, 2681–2694. https://doi.org/10.1007/s00415-012-6570-y.

Husted, J.A., Cook, R.J., Farewell, V.T., Gladman, D.D., 2000 May. Methods for assessing responsiveness: a critical review and recommendations. J. Clin. Epidemiol. 53 (5), 459–468. https://doi.org/10.1016/S0895-4356(99)00206-1.

Kamper, S.J., Maher, C.G., Mackay, G., 2009. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J. Man. Manip. Ther. 17 (3), 163–170. https://doi.org/10.1179/jmt.2009.17.3.163.

Kocher, M.S., Horan, M.P., Briggs, K.K., Richardson, T.R., O'Holleran, J., Hawkins, R.J., 2005 Sep. Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. J Bone Joint Surg Am 87 (9), 2006–2011. https://doi.org/10.2106/JBJS.C.01624.

Kon, S.S., Canavan, J.L., Jones, S.E., et al., 2014. Minimum clinically important difference for the COPD assessment test: a prospective analysis. Lancet Respir Med 2 (3), 195–203. https://doi.org/10.1016/S2213-2600(14)70001-3.

Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15 (2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

Luime, J.J., Koes, B.W., Hendriksen, I.J., et al., 2004. Prevalence and incidence of shoulder pain in the general population; a systematic review. Scand. J. Rheumatol. 33 (2), 73–81. https://doi.org/10.1080/03009740310004667.

McHugh ML, 2012. Interrater reliability: the kappa statistic. Biochem. Med. 22 (3), 276–282.

Michener, L.A., McClure, P.W., Sennett, B.J., 2002 Nov-Dec. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self report section: reliability, validity, and responsiveness. J. Shoulder Elb. Surg. 11 (6), 587–594. https://doi.org/10.1067/mse.2002.127096.

Minagawa, H., Yamamoto, N., Abe, H., et al., 2013. Prevalence of symptomatic and asymptomatic rotator cuff tears in the general population: from mass-screening in one village. J. Orthop. 10 (1), 8–12. https://doi.org/10.1016/j.jor.2013.01.008.

Mintken, P.E., Glynn, P., Cleland, J.A., 2009 Nov-Dec. Psychometric properties of the shortened disabilities of the Arm, shoulder, and Hand questionnaire (QuickDASH) and numeric pain rating scale in patients with shoulder pain. J. Shoulder Elb. Surg. 18 (6), 920–926.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., et al., 2010 May. 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. 19 (4), 539–549. https://doi.org/10.1007/s11136-010-9606-8.

Moser, A.D., Knaut, L.A., Zotz, T.G., Scharan, K.O., 2012 May-Jun. Validity and reliability

of the Portuguese version of the American shoulder and Elbow Surgeons standardized shoulder assessment form. Rev. Bras. Reumatol. 52 (3), 348–356. https://doi.org/10.1590/S0482-50042012000300005.

Odetunde, M.O., Akinpelu, A.O., Odole, A.C., 2018 Dec 20. Cross-cultural adaptation and validation of the stroke specific quality of life 2.0 scale into Hausa language. J Patient Rep Outcomes 2 (1), 63. https://doi.org/10.1186/s41687-018-0082-1.

Padua, R., Padua, L., Ceccarelli, E., Bondi, R., Alviti, F., Castagna, A., 2010 May. Italian versión of ASES questionnaire for shoulder assessment: cross-cultural adaptation and validation. Musculoskelet Surg 94 (1), S85–S90. https://doi.org/10.1007/s12306-010-0064-9.

Piitulainen, K., Paloneva, J., Ylinen, J., Kautiainen, H., Häkkinen, A., 2014. Reliability and validity of the Finnish version of the American shoulder and Elbow Surgeons standardized shoulder assessment form, patient self-report section. BMC Muscoskelet. Disord. 15, 272. https://doi.org/10.1186/1471-2474-15-272.

Post MW, 2016 Jul. What to do with "moderate"reliability and validity coefficients? Arch. Phys. Med. Rehabil. 97 (7), 1051–1052. https://doi.org/10.1016/j.apmr.2016.04.001.

Prinsen, C.A.C., Mokkink, L.B., Bouter, L.M., Alonso, J., Patrick, D.L., de Vet, H.C.W., Terwee, C.B., 2018 May. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual. Life Res. 27 (5), 1147–1157. https://doi.org/10.1007/s11136-018-1798-3.

Reed, M.D., Van Nostran, W., 2014 Mar. Assessing pain intensity with the visual analog scale: a plea for uniformity. J. Clin. Pharmacol. 54 (3), 241–244. https://doi.org/10.1002/jcph.250.

Revicki, D., Hays, R.D., Cella, D., Sloan, J., 2008 Feb. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J. Clin. Epidemiol. 61 (2), 102–109. https://doi.org/10.1016/j.jclinepi.2007.03.012.

Richards, R.R., An, K.N., LU, Bigliani, et al., 1994 Nov. A standardized method for the assessment of shoulder function. J. Shoulder Elb. Surg. 3 (6), 347–352. https://doi.org/10.1016/S1058-2746(09)80019-0.

Schmidt, S., Ferrer, M., González, M., et al., 2014 Mar. Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. J. Shoulder Elb. Surg. 23 (3), 434–444. https://doi.org/10.1016/j.jse.2013.09.029.

Tashjian, R.Z., Deloach, J., Green, A., Porucznik, C.A., Powell, A.P., 2010 Feb. Minimal clinically important differences in ASES and simple shoulder test scores after non-operative treatment of rotator cuff disease. J Bone Joint Surg Am 92 (2), 296–303. https://doi.org/10.2106/JBJS.H.01296.

Tashjian, R.Z., Hung, M., Keener, J.D., el al, 2017 Jan. Determining the minimal clinically important difference for the American Shoulder and Elbow Surgeons score, Simple Shoulder Test, and visual analog scale (VAS) measuring pain after shoulder arthroplasty. J. Shoulder Elb. Surg. 26 (1), 144–148. https://doi.org/10.1016/j.jse.2016.06.007.

Tavakol, M., Dennick, R., 2011 Jun 27. Making sense of Cronbach's alpha. Int. J. Med. Educ. 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd.

Terwee, C.B., Bot, S.D., de Boer, M.R., et al., 2007 Jan. Quality criteria were proposed for measurement properties of health status questionnaires. J. Clin. Epidemiol. 60 (1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.

Terwee, C.B., Mokkink, L.B., Knol, D.L., Ostelo, R.W.J.G., Bouter, L.M., de Vet, H.C.W., 2012. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual. Life Res. 21, 651–657. https://doi.org/10.1007/s11136-011-9960-1.

Vrotsou, K., Cuéllar, R., Silió, F., el al, 2016 Oct 18. Patient self-report section of the ASES questionnaire: a Spanish validation study using classical test theory and the Rasch model. Health Qual. Life Outcomes 14 (1), 147. https://doi.org/10.1186/s12955-016-0552-1.

Ware, J.E., Keller, S.D., Gandek, B., Brazier, J.E., Sullivan, M., 1995. Evaluating translations of health status questionnaires: methods from the IQOLA project. Int. J. Technol. Assess. Health Care 11 (03), 525–551. Summer. https://doi.org/10.1017/s0266462300008710.

Werner, B.C., Chang, B., Nguyen, J.T., Dines, D.M., Gulotta, L.V., 2016 Dec. What change in American Shoulder and Elbow Surgeons Score represents a clinically important change after shoulder arthroplasty? Clin. Orthop. Relat. Res. 474 (12), 2672–16781. https://doi.org/10.1007/s11999-016-4968-z.

Williamson, A., Hoggart, B., 2005 Aug. Pain: a review of three commonly used pain rating scales. J. Clin. Nurs. 14 (7), 798–804. https://doi.org/10.1111/j.1365-2702.2005.01121.x.

Wright, A., Hannon, J., Hegedus, E.J., Kavchak, A.E., 2012 Aug. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). J. Man. Manip. Ther. 20 (3), 160–166.

Yahia, A., Guermazi, M., Khmekhem, M., Ghroubi, S., Ayedi, K., Elleuch, M.H., 2011 Mar. Translation into Arabic and validation of the ASES index in assessment of shoulder disabilities. Ann Phys Rehabil Med 54 (2), 59–72. https://doi.org/10.1016/j.rehab.2010.12.002.

Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3 (1), 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.