Taylor & Francis
Taylor & Francis Group

Disability and Rehabilitation

ISSN: 0963-8288 (Print) 1464-5165 (Online) Journal homepage: www.tandfonline.com/journals/idre20

The reliability and validity of the Turkish version of the oxford shoulder instability score

Emel Sonmezer, Hayri Baran Yosmaoglu & Celal Deha Doğan

To cite this article: Emel Sonmezer, Hayri Baran Yosmaoglu & Celal Deha Doğan (2020) The reliability and validity of the Turkish version of the oxford shoulder instability score, Disability and Rehabilitation, 42:2, 261-266, DOI: 10.1080/09638288.2018.1497717

To link to this article: https://doi.org/10.1080/09638288.2018.1497717

DISABILITY AND REHABILITATION
2020, VOL. 42, NO. 2, 261–266
https://doi.org/10.1080/09638288.2018.1497717

Taylor & Francis
Taylor & Francis Group

ASSESSMENT PROCEDURE

The reliability and validity of the Turkish version of the oxford shoulder instability score

Emel Sonmezer[a], Hayri Baran Yosmaoglu[a] and Celal Deha Doğan[b]

[a]Department of Physical Therapy and Rehabilitation, Faculty of Health Sciences, Baskent University, Baglica/Ankara, Turkey; [b]Department of Measurement and Evaluation, Faculty of Educational Sciences, Ankara University, Cebeci/Ankara, Turkey

ABSTRACT

Background: The aim of this study was to adapt the Oxford Shoulder Instability Score to Turkish culture and test its reliability and validity.

Methods: This study included 118 patients with shoulder instability. Confirmatory factor analyses, and correlation coefficient between Oxford Shoulder Instability Score and Short Form 36 were calculated in order to test construct validity. Internal consistency was tested using Cronbach's alpha. Pearson correlation were calculated to test reliability. Differential item functioning analysis was performed to detect whether items exhibited differences according to gender.

Results: Confirmatory factor analysis indicating the single structure of the Oxford Shoulder Instability Score was confirmed. Cronbach's alpha was calculated as 0.87 for the whole scale. There were positive and strong correlations between the first and follow-up assessments ($r = 0.86$, $p < 0.01$). The Turkish version of OSIS showed moderate and significant correlations with domains of the SF-36 in general. Results also showed that there was no item exhibiting differential item functioning analysis in the Turkish version of Oxford Shoulder Instability Score.

Conclusion: The Turkish version of the Oxford Shoulder Instability Score is a reliable, valid, reproducible and practical tool. It can be used for patients with shoulder disorders and is recommended for clinical use.

ARTICLE HISTORY
Received 6 November 2017
Revised 2 July 2018
Accepted 3 July 2018

KEYWORDS
Health-related quality of life; health status questionnaires; transcultural adaptation; functional movement

➤ IMPLICATIONS FOR REHABILITATION

- Oxford Shoulder Instability Score is a 12-item tool measuring health-related quality of life and shoulder functions.
- Since the Turkish version of the Oxford Shoulder Instability Score is confirm to be a reliable, valid, and reproducible tool, it can be used in clinics to assess the functional status in patient with shoulder instability.
- It can be recommended to identify improvements in patients with shoulder problems for research purposes as well.

Introduction

Outcome measures are important for defining the functional abilities and life quality of patients [1,2]. A profound examination of outcome measures provides reliable decisions regarding treatment. Moreover, outcome measures determine the baseline function of the patient at the beginning of treatment. Thus, the progress of the treatment can be followed up. Outcome measures can be achieved by various methods. One of the most frequently used is questionnaires. Questionnaires are tools to assess the patient's own feelings about the symptoms of his/her illness [3].

Shoulder joint disorders are very common orthopedic problems around the world [4,5]. The rehabilitation process for an injured shoulder may take a long time. Therefore, following up the progress of rehabilitation with questionnaires is essential to establish a proper rehabilitation program. Dawson and colleagues [6] developed the 12-item Oxford Shoulder Instability Score (OSIS) questionnaire to measure the quality of life of patients with instability of the shoulder joint. OSIS consists of one major component with 12 items. Each item is scored from 0 to 5. A higher score indicates a worse shoulder condition. OSIS is a valid, reliable scale that can be used to assess the functional status of patients with shoulder instability [6].

Transcultural adaptation of OSIS has been done for many languages [7,8]. Adaptation of OSIS into the Turkish language has not been done yet. Thus, the purpose of this study was to adapt OSIS to the Turkish culture and to test its reliability and validity.

Materials and methods

The original version of OSIS was adapted to Turkish in this study. Some detailed information about the sample, instruments, adaptation process and statistical analysis of the study are presented below.

Patients

This study included 118 (mean age 52 ± 14) patients with shoulder instability problem, 82 (69%) of whom were female and 36 (31%)

CONTACT Emel Sonmezer emelsonmezer@gmail.com Department of Physical Therapy and Rehabilitation, Faculty of Health Sciences, Baskent University Eskisehir Yolu 18 km, 06790, Baglıca/Ankara, Turkey

Table 1. Patient characteristics.

Characteristics	*n* (%)	Total
Gender		
Male	36 (31%)	118
Female	82 (69%)	
Age (year)		
19–30	13	118
31–50	31	
51 and 70	61	
70 and over	9	

of whom were male. Patients who were referred to the physiotherapy clinic with an initial diagnosis of shoulder instability were included in the study. The patients' diagnose were also confirmed using the physical examination tests [9] defined for shoulder instability, 62 (53%) of the patients were traumatic structural, 24 (20%) of the patients atraumatic structural, and 32 (27%) of the patients were habitual nonstructural or muscle patterning disorder based shoulder instability. All measurements were completed before the beginning of the physiotherapy treatment program to prevent the possible effect of treatment on the outcome of the scale. Table 1 presents some demographic information about the patients.

Approval was obtained from the Baskent University Clinical Researches Ethics Committee (KA 13/53–2013) and all patients consented to participate in the study. Informed consent was obtained from all individual participants included in the study.

Original version of the Oxford Shoulder Instability Questionnaire: OSIS was originally developed and validated by Dawson and colleagues in 1999 [6]. The original form of OSIS has 12 items on a single factor. Internal Consistency (Cronbach's alpha = 0.91), reliability (test-retest $r = 0.97$, $p < 0.001$) and the construct validity of original OSIS is high [6].

Translation and cross-cultural adaptation procedure

To adapt the original version of OSIS into Turkish, the following steps were followed [10]:

1. Conceptual definition: Each item of the original version of OSIS was examined to ensure they would be equivalent in Turkish. When it was ensured that all items had the same conceptual meanings in Turkish, we proceeded to the forward translation.
2. Forward translation: Forward translation was performed independently by two translators. One of the translators had a clinical and medical background relevant to shoulder rehabilitation. The other translator was a linguistics specialist and offered a translation that reflects the language used by the common population. Both translators specialized in medical translations.
3. Back translation: Next, the translated version of OSIS was back translated into English independently by another two translators. Those translators did not see the original English wording of OSIS. One of these translators had a clinical and medical background and was knowledgeable about shoulder problems. The other translator was a native English speaker.
4. Back translation review: A group of experts including the researchers reviewed the back translations against the source version to highlight any discrepancies in meaning or terminology.
5. Pilot testing: The translated OSIS was pilot tested with at least five patients who were suffering from shoulder problems. Each patient was asked during the face-to face interview to comment on the wording that was difficult to understand and to suggest alternative wording/phrasing.
6. Pilot testing review: A group of experts including the researchers reviewed the comments on the pilot testing report to highlight any discrepancies in meaning or terminology used. The final form of OSIS was then produced.

Analysis of data

The final version was validated according to COSMIN (The COnsensus-based Standards for the selection of health Measurement INstruments) international guidelines [11].

Sample size analysis

The studies suggest that sample size should be calculated as 5 or 10 participants per item for the researches including multivariable analyses [12,13]. OSIS is 12-item functional scale therefore, sample size calculated as at least 60–120 subject in this study in order to produce reliable result. 129 subject was asked to participate the study. nine subjects did not attend to measurements. Study was completed with 118 subject with shoulder instability.

Internal consistency

Cronbach's alpha internal consistency coefficient was calculated for OSIS in order to explore the degree of the interrelatedness among the items in the Turkish version of OSIS.

Reliability

To assess the test–retest reliability of the Turkish version of OSIS, 118 patients from the first assessment were asked to complete and return a second questionnaire 48 h after. The Pearson correlation coefficient was calculated to explore the relationship between the two assessments. Moreover, repeated measures *t*-tests were performed to determine if there was a change in the distribution of the scores between the two tests.

Validity

Construct Validity: SF-36 domains were also used to test the construct validity of the Turkish version of OSIS. A 36-item short-form (SF-36) was constructed to survey health status in the clinic was developed by Ware and Sherbourne [14]. The SF-36 includes one multi-item scale that assesses eight health concepts: (1) limitations in physical activities because of health problems; (2) limitations in social activities because of physical or emotional problems; (3) limitations in usual role activities because of physical health problems; (4) bodily pain; (5) general mental health (psychological distress and well-being); (6) limitations in usual role activities because of emotional problems; (7) vitality (energy and fatigue); and (8) general health perceptions. SF-36 was adapted to Turkish by Kocyigit [15]. The Pearson correlation coefficient and Spearman's correlation coefficient were calculated to assess the relationship between the Turkish version of OSIS and the SF-36 domains. The Pearson correlation coefficient requires all variable scores to be continuous and normally distributed. Thus, the normality assumption was tested for all variables.

To define the construct validity of the Turkish version of OSIS, confirmatory factor analyses were performed as well. The Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of sphericity were calculated to test if the data were suitable for factor analysis. The Kaiser–Meyer–Olkin measure of sampling adequacy test and Bartlett's test are two indicators of the strength of the relationship among variables [12]. Bartlett's test examines if the population matrix resembles an identity matrix. After the confirmatory factor analysis, the model fit was assessed using the ratio of chi-squared to degrees of freedom (χ^2/sd), the goodness of fit index, the adjusted goodness of fit index, the comparative fit index, the non-normed fit index, the root mean square error of

Table 2. Fit indices and cutoff points [11,12].

Fit index	Cutoff point
χ^2/sd	≤2.5 excellent fit ≤5 mediocre fit
The goodness of fit index - The adjusted goodness of fit index - The comparative fit index - The non-normed fit index	≥0.95 excellent fit ≥0.90 good fit
The standard root mean square residual - The root mean square error of approximation	≤0.05 excellent fit ≤0.08 good fit ≤0.010 poor fit

approximation, and the standard root mean square residual. Table 2 presents the cutoff points for these fit indices [16,17].

Criterion Validity: Measurement invariance plays a very important role in interpreting test scores appropriately for individuals from different populations (e.g. gender, race) [18]. One way of assessing measurement invariance is to examine whether the items function differentially between the groups of interest. Differential item functioning analysis is used to identify test items that are differentially difficult for respondents who have the same level of knowledge, skill or ability, but differ in ways that should be irrelevant to their performance on the test [19]. An item exhibits differential item functioning if the probability of answering an item correctly or responding to a particular category differs for individuals from different groups with the same level of proficiency [20]. Differential item functioning statistics may be based on a latent model or observed score. In this study, differential item functioning statistics were calculated based on the observed score according to gender. Males were defined as the reference group and females were defined as the focal group. To detect items exhibiting differential item functioning in the Turkish version of OSIS, the Mantel chi-squared [21], standardized Liu-Agresti cumulative log-odds ratio and standardized Cox's noncentrality parameter were calculated.

Mantel Chi-Square test is distributed as a chi-squared with one degree of freedom (Mantel, 1963). Values greater than 3.84 (for a type-1 error probability of 0.05) shows that the item exhibits differential item functioning. Standardized Liu-Agresti cumulative log-odds ratio is the Liu-Agresti cumulative common log-odds ratio divided by the estimated standard error [22]. Standardized Cox's noncentrality parameter is Cox's noncentrality parameter estimator divided by the estimated standard error [23]. Standardized Liu-Agresti cumulative log-odds ratio and standardized Cox's noncentrality parameter values greater than 2 and less than -2 indicate the items exhibiting differential item functioning. DIFAS 5.0 software was used for the differential item functioning analysis.

Item discrimination: The discrimination levels of the Turkish version of OSIS items were computed to investigate whether there was a significant difference between the total scale scores and the factor scores of the participants from the upper and lower 27% Upper and lower 27% technique is commonly used to test discrimination index of both dichotomous and polytomous items. [24]. In this study, the discrimination levels of the scale items were calculated investigating whether there was a significant difference between scale scores of the upper and lower 27% of the participants. Item discrimination was calculated for the domains of OSIS. Independent sample t-tests were performed to determine if there was a significant difference between the scores of the participants from the upper and lower groups.

Results

Internal consistency

The internal consistency of the Turkish version of OSIS was found to be 0.87. This value indicates high internal consistency. All items correlated with a total score of more than 0.5, except items 1 and 5. In each case, Cronbach's alpha value was more above 0.85.

Table 3. Latent and observed variables.

Latent variable	Observed variables
Shoulder instability	M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12

Reliability

The results show that there was positive and strong correlation between the first and follow-up assessments ($r = 0.86$ $p < 0.01$). Moreover, the repeated measures *t*-test was calculated to determine if there was a change in the distribution of the scores between the two tests. The results show that the mean scores of the first and second assessments were not significantly different ($t_{(117)}$=0.204 $p > 0.05$).

Validity

Construct validity: The results show that the Turkish version of OSIS and its dimensions had moderate to low significant correlations with most of the domains of SF-36. There was no significant correlation between the Turkish version of OSIS and the general health, mental health and vitality domains of SF-36. Table 4 presents the correlation between all SF-36 domains and the Turkish version of the OSIS domains.

According to the results of the Kaiser–Meyer–Olkin measure of sampling adequacy test, the Kaiser–Meyer–Olkin measure of sampling adequacy value was 0.856. Bartlett's test was found to be highly significant ($p < 0.001$). Considering the results of the The Kaiser–Meyer–Olkin measure of sampling adequacy test and Bartlett's test, the data were found to be suitable for factor analysis.

The original form of OSIS has 12 items on a single factor. Table 3 presents the observed variables, latent variables and abbreviations for latent variables according to confirmatory factor analysis.

The results of the confirmatory factor analysis show that the *t* values for all the relations between the latent and observed variables were above the critical ratio (2.99) and are statistically significant at the 0.01 level.

Figures 1 and 2 demonstrate the *t* values and standardized coefficients, respectively.

The modification indices show that there was a significant decrease in the χ^2 value when items 3–10 and 9–12 were modified. Based on reviews and expert views, it was decided that the modifications had a theoretical basis. Following modifications, the analysis was re-conducted and the ratio of χ^2 to degrees of freedom was calculated as 132.53\53 = 2.55. This result indicates an excellent fit [17]. The results of the other fit indices were as follows: the non-normed fit index: 0.92, the comparative fit index: 0.94, the goodness of fit index: 0.84, the adjusted goodness of fit index: 0.76, the root mean square error of approximation: 0.10 and standardized root mean square residuals: 0.07. Because the comparative fit index and the non-normed fit index values were above 0.95, this indicates a good fit [10,20]. The standardized root mean square residuals value was below 0.08 and also indicated a

good fit. The goodness of fit index value indicated a mediocre fit. The root mean square error of approximation value showed poor fit as it was below 0.010 and the adjusted goodness of fit index values indicated a poor fit [17,25].

These findings indicate that the ratio of χ^2 to degrees of freedom and the non-normed fit index, the comparative fit index, the goodness of fit index and standardized root mean square residuals values were at the expected level and indicated a good to mediocre fit. Although the adjusted goodness of fit index value was a bit lower and the root mean square error of approximation value was a bit higher than expected, the single factor structure of OSIS was confirmed.

Criterion Validity: No items in the Turkish version of OSIS exhibited differential item functioning according to gender. These findings suggest that the items of OSIS are not biased regarding the gender of the patient. The differential item functioning results according to Mantel chi-squared, Standardized Liu-Agresti cumulative log-odds ratio and standardized Cox's non-centrality parameter are presented in Table 5.

As shown in Table 5, the Mantel values for all items varied between 0.004 and 3.53. There was no item greater than the critical value of 3.84. Thus, no items exhibit differential item functioning according to the Mantel chi-squared test. According to the standardized Liu-Agresti cumulative log-odds ratio results, no items exhibiting differential item functioning were detected, since there were no values greater than 2 or less than -2 (standardized Liu-Agresti cumulative log-odds ratio values for all items ranged between -1.205 and 1.745). The standardized Cox's non-centrality parameter values ranged between 1.87 and 1.17; thus, no item exhibited differential item functioning since no values were greater than 2 or less than −2.

Item discrimination: The test results indicate that the mean of each item of the Turkish version of OSIS was higher for the upper 27%; this difference was significant at the 0.01 level. This result shows that the items of OSIS could discriminate between participants in the lower and upper 27% (Mean ± SD for upper 27%, lower 27%, and remaining groups: 45.5 ± 13.29, 27.22 ± 2.44, and 36.3 ± 3.09, respectively). In other words, the discrimination level of the Turkish version of OSIS was high.

Table 4. Correlation between SF-36 and OSIS domains.

MOS SF-36	Correlation with Turkish version of OSIS	Correlation in original version of OSIS*
Physical functioning	−0.34**	−0.71**
Social functioning	−0.32**	−0.58**
Pain	−0.59**	−0.70**
Role physical	−0.54**	−0.69**
Role emotional	−0.37**	−0.30**
Vitality	−0.12	−0.20
Mental health	−0.14	−0.35**
General health	−0.02	−0.08

*The data were taken from the validation study of original version of OSIS [6].
**$p < 0.01$.

Discussion

Main findings of this study suggested that the Turkish version of OSIS was reliable. The correlation between the pre- and post-tests was high. Additionally, there was no significant difference between the means of the pre- and post-tests. These results are consistent with the values for the original version of OSIS, which was found to be a valid and reliable tool to assess the functional status of a patient with shoulder instability.

The findings of our study also suggested that the Turkish version of the OSIS is valid. The construct validity results also identified that the Turkish version of OSIS has moderate and significant correlations with the domains of the SF-36 in general. There was an insignificant correlation between the Turkish OSIS version and the general health and vitality and mental health domains of the SF-36. This finding is also consistent with the original study describing the development of OSIS in that there was also a

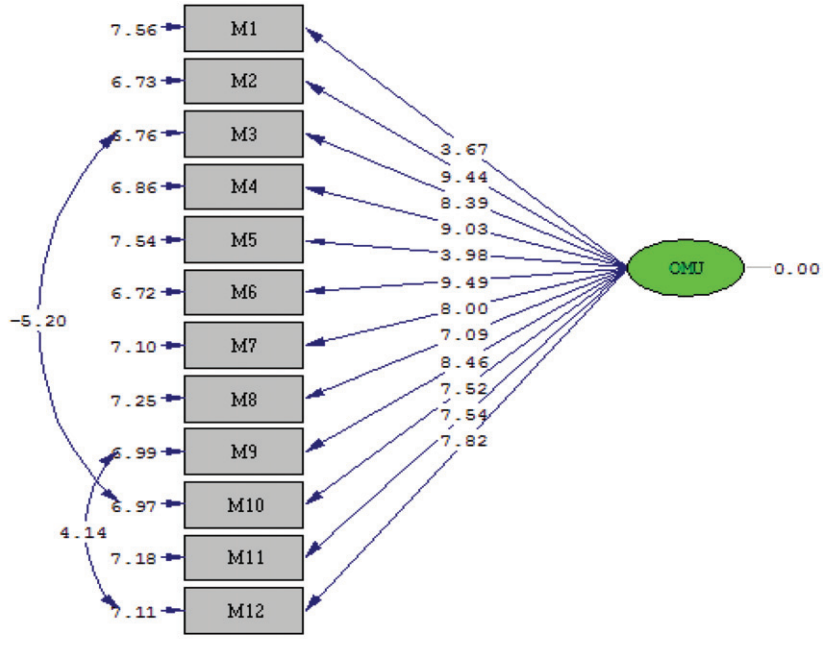

Figure 1. *t* values for the relation between latent and observed variables.

nonsignificant correlation between the original OSIS and the general health and vitality domains of SF-36. There was significant but weak correlation between the original version of OSIS and the mental health domain of SF-36 [6]. Similarly, in previous study, Dutch version of OSIS was more closely correlated with the SF-36 subscales "pain" (0.78) and "role physical" (0.69) than with the subscale "physical function" (0.65) [8]. Similar to our results, high test–retest reliability results (ICC: 0.87) were found in Dutch version of OSIS as well [8].

Although the root mean square error of approximation and the adjusted goodness of fit index values were not at the expected level, the ratio of χ^2 to degrees of freedom, the non-normed fit index, the comparative fit index, the goodness of fit index and standardized root mean square residuals values were at the expected level, indicating an excellent to good fit according to confirmatory factor analysis. The original form of the OSIS has 12 items in a single factor and the single structure of the Turkish version of OSIS was confirmed as well.

Findings of the presented study suggested that the internal consistency of the Turkish version of the OSIS was high (Cronbach's $\alpha = 0.87$). This value is very close to the value presented in the original version of the OSIS and Dutch version of OSIS (Cronbach's $\alpha = 0.91$, 0.88, respectively) [6,8].

The findings of our study also show that the mean of each item on the scale was higher for the upper 27%. This indicates that items in the Turkish version of OSIS discriminate between patients who have significant shoulder problems (upper 27%) and those who have minor shoulder problems (lower 27%). This indicates that the Turkish version of OSIS could be used by clinicians to determine the severity of shoulder problems.

The Mantel chi-squared, standardized Liu-Agresti cumulative log-odds ratio and standardized Cox's noncentrality parameter results show that the items of the Turkish version of OSIS do not exhibit differential item functioning according to the gender of the patient. Therefore, it can be said that the Turkish version of OSIS is at the expected level in terms of measurement invariance. In other words, there is no item that is biased according to gender in the Turkish version of OSIS.

Table 5. DIF analysis for the Turkish version of OSIS.

ITEM	MANTEL	Standardized Liu-Agresti cumulative log-odds ratio	Standardized Cox's noncentrality parameter
Item1	0.06	−0.26	−0.25
Item2	0.66	0.82	0.81
Item3	3.53	−1.72	−1.88
Item4	1.38	−1.20	−1.18
Item5	0.94	−1.01	−0.97
Item 6	0.00	−0.07	−0.07
Item 7	2.16	1.47	1.47
Item 8	0.42	−0.61	−0.65
Item 9	2.9	1.75	1.72
Item 10	0.75	0.90	0.87
Item 11	2.95	−1.69	−1.72
Item 12	1.37	1.31	1.17

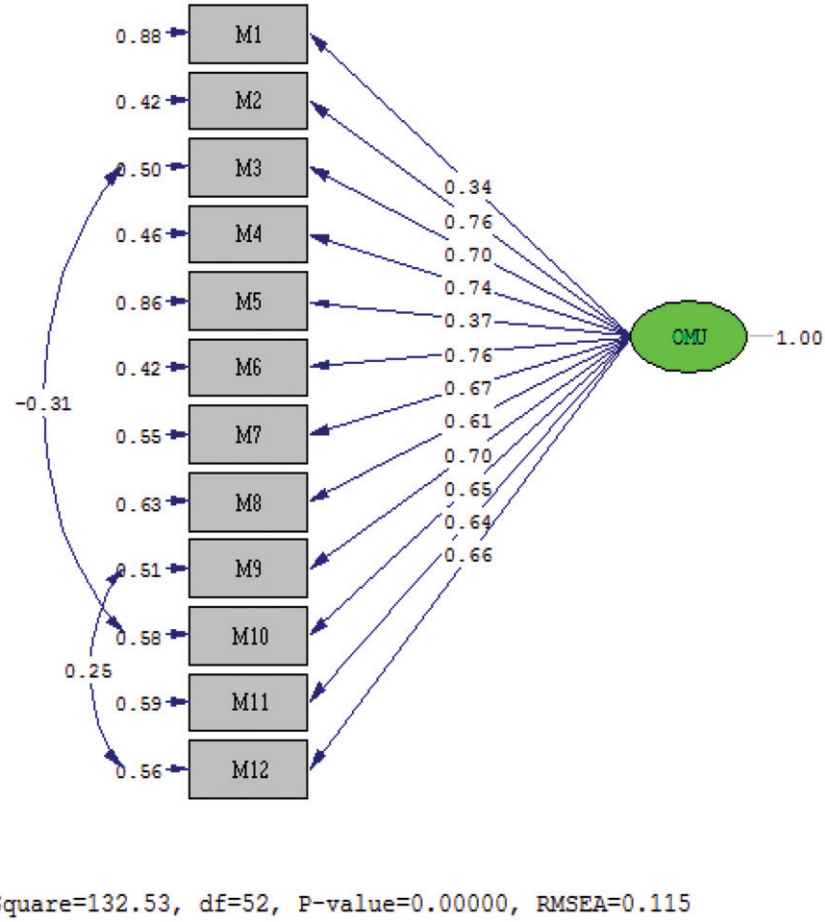

Figure 2. Standardized coefficients for the relation between latent and observed variables.

Conclusion

The Turkish version of OSIS has high internal consistency, is reproducible, can discriminate between patients with significant and minor shoulder problems, has high construct validity and has a moderate correlation with SF36 domains. The Turkish version of OSIS can be used to assess the functional status of patients for research purposes. The Turkish version of OSIS can also be used to identify improvements in patients with shoulder problems in the clinic.

Disclosure statement

The authors report no declarations of interest.

References

[1] Fitzpatrick R, Fletcher A, Gore S, et al. Quality of life measures in health care. I: Applications and issues in assessment. BMJ. 1992;305:1074–1077.
[2] Testa MA, Simonson DC. Assessment of quality-of-life outcomes. N Engl J Med. 1996;334:835–840.
[3] Patrick D, Guyatt G, Acquadro C. Patient-reported outcomes. In: J Higgens and S Green, editors. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons; 2008. p. 531–545.
[4] Robinson CM, Howes J, Murdoch H, et al. Functional outcome and risk of recurrent instability after primary traumatic anterior shoulder dislocation in young patients. J Bone Joint Surg Am. 2006;88:2326–2336.
[5] Leroux T, Ogilvie-Harris D, Veillette C, et al. The epidemiology of primary anterior shoulder dislocations in patients aged 10 to 16 years. Am J Sports Med. 2015;43:2111–2117.
[6] Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability. The development and validation of a questionnaire. J Bone Joint Surg Br. 1999;81:420–426.
[7] Olyaei G, Mousavi S, Montazeri A, et al. Translation and Validation Study of the Persian Version of the Oxford Shoulder Instability Score. J Modern Rehabil. 2016; 10:23–27.
[8] van der Linde JA, van Kampen DA, van Beers LW, et al. The Oxford Shoulder Instability Score; validation in Dutch and first-time assessment of its smallest detectable change. J Orthop Surg Res. 2015;10:146.
[9] Gerber C, Nyffeler RW. Classification of glenohumeral joint instability. Clin Orthop Related Res. 2002;400:65–76.
[10] 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 Health. 2005;8:94–104.
[11] Mokkink LB, Prinsen CAC, Bouter LM, et al. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther. 2016; 20:105–113.
[12] Field A. Discovering statistics Using SPSS. London (UK): Sage Publication; 2005.
[13] Kline P. An easy guide to factor analysis: New York (NY): Routledge; 1994.
[14] Ware JE, Sherbourne DC. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection). Med Care. 1992;30:473–483.
[15] Bilir Kaya B, İçağasıoğlu A. Reliability and validity of the Turkish version of short form-36 (SF-36): a study in a group of patients will rheumatic diseases. Turk J Drugs Ther. 2018;2:102–106.
[16] Sümer N. Yapısal eşitlik modelleri. Türk Psikoloji Yazıları 2000;3:49–74.
[17] Tabachnick BG, Fidell LS. Using multivariate statistics. California: Pearson; 2001.
[18] Raju NS, Laffitte LJ, Byrne BM. Measurement equivalence: a comparison of confirmatory factor analysis and item response theory. J Appl Psychol. 2002;87: 517–529.
[19] Miller T, Chahine S, Childs RA. Detecting differential item functioning and differential step functioning due to differences that should matter. Pract Assess Res Eval. 2010;15:1–13.
[20] Elosua P, Wells CS. Detecting DIF in polytomous items using MACS, IRT and ordinal logistic regression. Psicológica. 2013;34:327–342.
[21] Mantel N. Chi-square tests with one degree of freedom: Extension of the Mantel-Haenszel procedure. J Am Stat Assoc. 1963;58:690–700.
[22] Penfield RD, Algina J. Applying the Liu-Agresti estimator of the cumulative common odds ratio to DIF detection in polytomous items. J Educ Measure. 2003;40: 353–370.
[23] Camilli G, Congdon P. Application of a method of estimating DIF for polytomous test items. J Educ Behav Stat. 1999;24:323–341.
[24] Crocker LM, Algina J. Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston; 1986.
[25] Hooper D, Coughan J, Mullen M. Structural equation modeling: Guidelines for determining model fit. Electron J Business Res Methods. 2008;6:53–60.