



# Validity and reliability of the Arabic version of the Back Pain Function Scale in patients with low back pain

Doaa Rafat ElAzab<sup>1</sup>, Ahmed Mahmoud Kadry<sup>2</sup>, Eman .A. Elhosary<sup>3</sup>

*1 Basic science Department, Faculty of Physical therapy, Cairo University, Cairo, Egypt*

*2 Integumentary Department; Faculty of Physical therapy, Kafr Elshiekh University, Kafr Elshiekh, Egypt*

*3 Women Health Department, Faculty of Physical therapy, Kafr Elshiekh University, Kafr Elshiekh, Egypt*

## \*Correspondence to

Eman. A. Elhosary, Department of Women Health, Faculty of Physical therapy, Kafr Elshiekh University, Kafr Elshiekh, Egypt

## Published online:

June 2021

## Abstract:

**Purpose:** the translation, cross-cultural adaptation and validation of the Back Pain Functional Scale (BPFS) into Arabic.

**Methods:** According to standardized guidelines, the translation and adaptation of the scale was completed. Two panels of experts were asked to review the instrument for consistency and to evaluate all the elements of the scale for content relevance. A convenient sample of 65 patients complained of LBP, completed BPFS and recompleted it to test-retest reliability after 2 days.

**Results:** The Arabic adapted version of the BPFS was produced and validated. The scale index of clarity was 99%. Index of Content Validity (ICV) = 98.3%. The Cronbach's  $\alpha$  coefficient for internal consistency was 0.865. The Total score of Spearman's correlation coefficients is 0.878.

**Conclusion:** The BPFS presented excellent reliability and validity for assessing the functional status in Arabic speaking patients with LBP.

**Key words:** Validity; reliability; Back Pain Function Scale; low back pain

## 1.Introduction

Low back pain (LBP) is a very prevalent global health concern and a major cause of disability that affects job efficiency and overall well-being (1). Visiting doctors is a very common explanation (2, 3). LBP affects and induces pain, disability, and psychosocial issues for people of all ages. LBP causes economic burden and huge direct and indirect costs (4)

It is a benign case, and in most situations it resolves naturally (5,6,7). It affects active young adults selectively. While clinical evaluation is critical in assessing patients, the effect of the disease on the functional state of the patient may not be assessed (6, 8, 9) Several scales are available for the functional assessment of low back pain patients, but few of which have been validated. (10,11,12).

In patients with LBP, the assessment of activity limitation and participation restriction is an essential component and patient-reported outcome measures

(PROMs) tend to be superior to standard physiotherapeutic disability measures such as range of motion in assessing patient outcomes. (13).

Outcome measures can show disability, limitation of activity and restriction of participation (14). Practicality, accuracy, reliability, validity, responsiveness and the ability to identify changes in a particular situation are essential characteristics of an outcome measure. (15).

Bias can be removed by self-administered questionnaires and, while subjective, it is considered to be a systematic, reproducible and responsive

clinical improvement. In order to use any new evaluation instrument in analysis, eliminating bias in the study should be valid and accurate (16).

The researcher is able to obtain similar data from different samples and from different backgrounds by adapting an instrument to a language other than the original, which allows greater fairness in the assessment. The use of adapted methods provides a greater ability to generalize findings and allows variations to be explored within an increasingly distinct community. (17).

The Back Pain Function Scale (BPFS) is one of the PROMs designed to overcome the limitations of current PROMs in clinical practice and clinical study settings to determine the functional state of individual patients with LBP (18). The BPFS has been shown to have sound reliability and validity measurement properties for assessing the functional status of LBP patients. (18, 19). The BPFS is in English, and before using it in Arabic-speaking countries, it must undergo cultural adaptation and validation. The goal of this study was, therefore, to translate and cross-culturally adapt the BPFS to the Arabic language and to test the validity and reliability of the Arabic version of BPFS in patients with LBP complaints who speak Arabic.

## 2. Patients and Methods

### 2.1. Study participants and recruitment criteria:

This study was conducted in the physical therapy clinic of Kafr Elsheikh University Hospital, Kafr Elsheikh city, Egypt, to translate, and cross culturally adapt the BPFS to Arabic language and to test the validity and reliability of the Arabic version of BPFS in Arabic-speaking patients with LBP complaints. A convenient sample of 65 patients complains from LBP were recruited from, Outpatient Clinic of Orthopedic Department. For their ability to participate in the research, they were enrolled and tested. They are able (not illiterate) to read and write, have not been deformed, and no prescriptions have been given. Patients fill in the data collection sheet used to gather demographic information (name, age, sex, weight, height, and BPFS). Also included in the study were two qualified panels, each containing 10 experts with at least 10 years of experience or a master's degree.

### 2.2. Study Design:

This is a prospective observational study for testing face and content validity, internal consistency and test-retest reliability of a translated scale.

Ethical approval was obtained from the institutional review board at Faculty of physical therapy, Kafr Elsheikh University before study commencement with number P.T./BAS/2/2021/4. Written informed

consent was obtained from all participants before the study.

### 2.3. Methods:

1) Forward translation: Primary scale translation from English to Arabic: Two translators were involved in forward translation. To generate two Arabic (Ar1 and Ar2) versions of the scale.

2) Preliminary development of the Arabic version. Both versions of the preliminary Arabic version of the scale (Ar1 and Ar2) were compared and combined by the researchers: blind back translation (blind revision translation or blind double revision translation). The backward translation was done by two translators. To produce two back-translated English revision versions, the preliminary translated Arabic version of the scale was translated into English (En1 and En2).

3) Comparison of the two back-translated English versions of the scale (En1 and En2): researchers compared the two English back translation versions of the scale (En1 versus En2) and compared both En1 and En2 with the original English scale in terms of commands, objects, format of answers, meanings of words, structure of sentences, importance, and relationships. No major variations were found between them; the preliminary translated Arabic version was therefore considered to be the pre-final Arabic version of the scale.

4) Pilot study to assess the intrarater reliability of the pre-final Arabic version of the scale: using a dichotomous scale (yes/no), the first expert panel was asked to evaluate each word (instructions, objects, and choices) of the tool for clarification. If the answer was no, they were asked to give suggestions for improving its clarity. Slight adjustments were made to increase the clarity index to the minimum appropriate value (80 percent) so that it could be filled by the patients, according to the recommendations of the first expert group. The second panel of experts was asked to rank and evaluate all material equivalence (relevance) scale items using a scale containing four grades: 1; mean not important, 2; mean unable to determine relevance, 3; mean relevant but requiring minor changes, and 4; mean very relevant and succinct. Suggestions for improving its relevance were then asked and the findings were interpreted as follows: the first and second grades were considered to be insignificant, while the third and fourth grades were considered to be relevant. It was named the final Arabic version of the BPFS after the pre-final version passed the clarification and relevance test.

5) The final Arabic version of the scale intrarater reliability test was performed on LBP patients as follows: patients completed the data collection sheet

used to collect demographic data (name, age and gender) and the final Arabic version of the BPFs.

6) After 2 days, the patients completed the scale again.

### 3. DATA ANALYSIS:

For data analysis, the IBM SPSS computer program ver. 23.0 was used. Clarity index and expert proportion of the clearance were used to test face validity. Index of content validity (CVI) and expert proportion of relevance were used to test the content validity. Cronbach's coefficient alpha was used to measure the internal consistency reliability. Test-retest reliability was measured using mean scores and Spearman's rank correlation.

### 4. RESULTS:

As shown in table (1) Study group consisted of 65 Patients 12 Male and 53 Female table (2), there mean age value was  $(41.57 \pm 11.43)$  years ranged from 30 to 53 years. Also, as shown in table 2 the mean weight of subjects is  $(76.36 \pm 15.32)$  and mean height of  $(168.5 \pm 10.78)$  while the mean of BMI is  $(31.7 \pm 3.85)$ .

**Table 1: Sex distribution in the study group**

Group	Study group	
Sex	Male	Female
Frequency	12	53

**Table 2: General characteristics of patients**

	Mean	Stander Deviation	Minimum	Maximum
Age	41.57	11.43	30	53
Weight	76.36	15.32	61.04	91.68
Height	168.5	10.78	157.7	179.3
BMI	31.7	3.85	27.8	35.6

The scale index of clarity was 93.3% of the prefinal Arabic version of the BPFs, as shown in Table 3, while the scale index of clarity was 99% of the final Arabic version of the BPFs, as shown in Table 4.

Content validity analysis: As shown in Table (5) according to the experts' opinions, 12 items were relevant with Index of Content Validity (ICV)= 98.3%,

So the content validity of the Arabic version of BPFs was perfect according to experts' opinions.

**Table 3: Item index of clarity of the pre final Arabic version of the scale**

Item number	Number of expert agree	Item index of clarity
1	10	100%
2	10	100%
3	9	90%
4	10	100%
5	8	80%
6	7	70%
7	10	100%
8	10	100%
9	8	80%
10	10	100%
11	10	100%
12	10	100%
Mean	9.33	93.3%

**Table 4: Item index of clarity of the final Arabic version of the scale**

Item number	Number of expert agree	Item index of clarity
1	10	100%
2	10	100%
3	10	90%
4	10	100%
5	10	100%
6	9	100%
7	10	100%
8	10	100%
9	10	100%
10	10	100%
11	10	100%
12	10	100%
Mean	9.9	99%

Internal consistency was calculated for the BPFSS final version. Cronbach's alpha was 0.865 with lower bound 0.798 and upper bound 0.912 at 95% confidence interval (CI). Finally, Spearman's correlation coefficients were calculated and are presented in Table 6.

**Table 5: Item index of content validity of the final Arabic version of the scale**

Item number	Number of expert agree (relevant responses)	CVI
1	10	100%
2	10	100%
3	10	90%
4	10	100%
5	10	100%
6	9	100%
7	10	100%
8	10	100%
9	9	100%
10	10	90%
11	10	100%
12	10	100%
Mean	9.83	98.3%

## 5.DISCUSSION:

In order to detect the natural evolution of diseases and the functional limitations of patients, clinicians need measures that are easy to use, accurate, reliable, and responsive to changes. Unless it is accurate and reproducible, a calculation is useless. Validity and reliability verify the accuracy of measurements (20).

The Arabic version of the BPFSS exhibited very good face validity, with scale index of clarity 93.3%. The scale also demonstrated excellent content validity because the ICV was 98.3%, the present results are in agreement with the reports of Polit and Beck (21) Who reported that it should be composed of elements with item indices of content validity (ICVI) that meet the following requirements (I-CVI of 1.00 for three to five experts and I-CVI .0.78 for 6.10 experts) for a scale to be judged to have excellent content validity and have S-CVI .0.90. Two rounds of expert analysis may be required for the prescribed requirements if the initial assessment shows the need for significant item changes.

This is consistent with the results of Waltz et al. (22), who noted that the minimum acceptable index is S-CVI/average .0.90, and updated and reevaluated items that do not reach the minimum acceptable indices.

The internal consistency reliability was excellent (Cronbach's  $\alpha$  of 0.865) similar to the original English BPFSS: Cronbach's  $\alpha = 0.93$  (18). Additionally, values of item-total correlation were appropriate. Furthermore, when an item was removed, the alpha values were comparable to the overall alpha coefficient of Cronbach, which implies that the BPFSS items are interrelated and that each item contributes similarly to the calculated build. Such results indicate that a single, latent construct is evaluated by the BPFSS. The reliability of the Persian edition of BPFSS is assured by a reasonably adequate Cronbach value comparable to the original English BPFSS. Therefore, according to Darren and Mallery (23), an  $\alpha$  value between 0.7 and 0.9 indicates good internal consistency.

The Arabic version of the BPFSS exhibited good test-retest reliability, with as all Spearman's rank correlation coefficients between test and retest results being statistically significant (item 1=0.913, item 2=0.875, item 3=0.895, item 4=0.935, item 5=0.815, item 6=0.764, item 7=0.874, item 8=0.922, item 9=0.789, item 10=0.927, item11=0.887, item 12=0.936, and total score=0.878). Therefore, according to Darren and Mallery (23), Spearman's rank correlation coefficient between 0.9 and 1 (as in items 1, 4, 8, 10, and 12) indicates excellent test-retest reliability, and Spearman's rank correlation coefficient between 0.7 and 0.9 (as in items 2, 3, 5, 6, 7, 9, and 11) indicates acceptable test-retest reliability.

Our results agree with Nakhostin Ansari., et al (24). They were studying the cross-cultural adaptation process of the original English BPFSS into the Persian language and testing the reliability of internal accuracy and the scale's test-retest reliability. The Cronbach's  $\alpha$  coefficient for internal consistency was 0.895. The test-retest reliability was excellent (ICC agreement = 0.88, CI 95%: 0.80–0.93). The standard error of measurement and smallest detectable change were 9.9% and 27.5%, respectively. Construct convergent validity was demonstrated with the Persian Functional Rating Index ( $r = -0.77$ ). Criterion validity was established with the Numerical Pain Rating Scale ( $r = -0.67$ ). They concluded that in Persian-speaking patients with LBP, the BPFSS in Persian language offered excellent reliability and validity for evaluating functional status.

**Table 6: Spearman's rank correlations coefficients**

Item number	R- value	Correlation strength	P value
1	0.913	Very strong	Statistically significant
2	0.875	Very strong	Statistically significant
3	0.895	Very strong	Statistically significant
4	0.935	Very strong	Statistically significant
5	0.815	Very strong	Statistically significant
6	0.764	strong	Statistically significant
7	0.874	Very strong	Statistically significant
8	0.922	Very strong	Statistically significant
9	0.789	strong	Statistically significant
10	0.927	Very strong	Statistically significant
11	0.887	Very strong	Statistically significant
12	0.936	Very strong	Statistically significant
<b>Total score</b>	<b>0.878</b>	<b>Very strong</b>	<b>Statistically significant</b>

## 5. CONCLUSION:

the present study successfully adapted the BPFS into Arabic language and presented acceptable reliability and validity consistent with the original English version. The validated BPFS facilitates international investigations using the BPFS as an outcome measure.

## References

1. Atlas SJ, Deyo RA. Evaluating and managing acute low back pain in the primary care setting. *Journal of general internal medicine.* 2001 Feb;16(2):120-31.
2. Deyo RA, Mirza SK, Martin BI. Back pain prevalence and visit rates: estimates from US national surveys, 2002. *Spine.* 2006 Nov 1;31(23):2724-7.
3. Jöud A, Petersson IF, Englund M. Low back pain: epidemiology of consultations. *Arthritis care & research.* 2012 Jul;64(7):1084-8.
4. Gore M, Sadosky A, Stacey BR, Tai KS, Leslie D. The burden of chronic low back pain: clinical comorbidities, treatment patterns, and health care costs in usual care settings. *Spine.* 2012 May 15;37(11): E668-77.
5. P. Pillastrini, I. Gardenghi, F. Bonetti, F. Capra, A. Guccione, et al. An updated overview of clinical guidelines for chronic low back pain management in primary care. *Joint Bone Spine,* 79 (2012), pp. 176-185
6. J.P. Valat. Épidémiologies des lombalgies. *Rev Rhum,* 65 (1998), pp. 172-174
7. J.P. Valat, S. Rozenberg, L. Bellaïche. Lombalgie. critères cliniques et d'imagerie. *Rev Rhum Monogr,* 77 (2010), pp. 158-166.
  - a. Beursksens, H. de Vet, A. Köke, G. van der Heijden, P. Knipschild. Measuring the functional status of patient with low back pain: assesment of the quality of four disease-specific questionnaires. *Spine,* 20 (1995), pp. 1017-1028.
8. Plénet, J. Gourmelen, J.F. Chastang, A. Ozguler, J. L. Lanoë, A. Leclerc. Seeking care for lower back pain in the French population aged from 30 to 69: the results of the 2002–2003 Décennale Santé survey. *Ann Phys Rehabil Med,* 53 (2010), pp. 224-231
9. Beursksens, H. de Vet, A. Köke. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain,* 65 (1996), pp. 71-76
10. P. Calmels, F. Béthoux, A. Condemine, I. Fayolle-Minon. Outils de mesure des paramètres fonctionnels dans la lombalgie. *Ann Reabil Med Phys,* 48 (2005), pp. 288-297
11. M. Davidson, J.L. Keating. A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther,* 82 (2002), pp. 8-24
12. Fairbank J, Pynsent P: The Oswestry Disability Index. *Literature Review.* *Spine.* (2000); vol. 25, no. 22, pp. 2940-2952.
13. Grimmer K, Bialocerkowski: A Survys. *Aust J Physiother.* (2005); vol. 51, no. 3, pp. 185-187.
14. Lurie J: A Review of Generic Health Status Measures in Patients with Low Back Pain. *Spine.* (2000); vol. 25, no. 24, pp. 3125-3129.
15. Kimberlin C, Winterstien A: Validity and Reliability of Measurement Instruments Used in Research. *American Journal of Health-System Pharmacists.* (2008);65(1),2276-2284.
16. Hambleton R: Issues, designs and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P.F. Merenda & C.D. Spielberger (Eds.). *Adapting educational and psychological tests for cross-cultural assessment.* Mahwah. NJ. Lawrence Erlbaum. (2005).
17. Stratford PW, Binkley JM, Riddle DL. Development and initial validation of the back pain

- functional scale. *Spine*. 2000 Aug 15;25(16):2095-102.
18. Longo UG, Loppini M, Denaro L, Maffulli N, Denaro V. Rating scales for low back pain. *British medical bulletin*. 2010 Jun 1;94(1):81-144.
  19. Giannini F, Cioni R, Mondelli M, Padua R, Gregori B, D'Amico P, Padua L. A new clinical scale of carpal tunnel syndrome: validation of the measurement and clinical-neurophysiological assessment. *Clinical neurophysiology*. 2002 Jan 1;113(1):71-7.
  20. Polit DF, Beck CT. The content validity index: are you sure you know what is being reported?: critique and recommendations. *Res Nurs Health* 2006; 29:489-97.
  21. Waltz CF, Strickland OL, Lenz ER. *Measurement in nursing and health research*. 3rd ed. New York (NY): Springer Publishing Company; 2005.
  22. Darren G, Mallery P. *SPSS for Windows step by step: a simple guide and reference 11.0 update*. 4th ed. Boston (MA): Allyn and Bacon; 2003.
  23. Nakhostin Ansari N, Naghdi S, Habibzadeh F, Salsabili N, Ebadi S. Persian translation and validation of the Back Pain Functional Scale. *Physiotherapy theory and practice*. 2018 Mar 4;34(3):223-30.