

RESEARCH ARTICLE

Translation into modern standard Arabic, cross-cultural adaptation and psychometric properties' evaluation of the Lower Extremity Functional Scale (LEFS) in Arabic-speaking athletes with Anterior Cruciate Ligament (ACL) injury

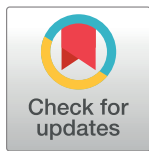
Vasileios Korakakis^{1,2,3}*, Michael Saretsky¹, Rodney Whiteley¹, Matthew C. Azzopardi^{1‡}, Jasenko Klauznicer^{1‡}, Abdallah Itani^{1‡}, Omar Al Sayrafi^{1‡}, Giannis Giakas^{2‡}, Nikolaos Malliaropoulos^{4,5,6‡}

1 Aspetar, Orthopaedic and Sports Medicine Hospital, Doha, Qatar, **2** Faculty of Physical Education and Sport Science, University of Thessaly, Trikala, Greece, **3** Hellenic Orthopaedic Manipulative Therapy Diploma (HOMTD), Athens, Greece, **4** Sports and Exercise Medicine Clinic, Thessaloniki, Greece, **5** National Track & Field Centre, Sports Medicine Clinic, Thessaloniki, Greece, **6** European Sports Care, London, United Kingdom

* These authors contributed equally to this work.

‡ These authors also contributed equally to this work.

* Vasileios.Korakakis@aspetar.com



OPEN ACCESS

Citation: Korakakis V, Saretsky M, Whiteley R, Azzopardi MC, Klauznicer J, Itani A, et al. (2019) Translation into modern standard Arabic, cross-cultural adaptation and psychometric properties' evaluation of the Lower Extremity Functional Scale (LEFS) in Arabic-speaking athletes with Anterior Cruciate Ligament (ACL) injury. PLoS ONE 14(6): e0217791. <https://doi.org/10.1371/journal.pone.0217791>

Editor: William Marshall Land, University of Texas at San Antonio, UNITED STATES

Received: January 5, 2019

Accepted: May 18, 2019

Published: June 10, 2019

Copyright: © 2019 Korakakis et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant, anonymized data are within the paper and its Supporting Information files.

Funding: The publication of this article was funded by the Qatar National Library.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Background

The Lower Extremity Functional Scale evaluates the functional status of patients that have lower extremity conditions of musculoskeletal origin. Regional Arabic dialects often create barriers to clear communication and comparative research. We aimed to cross-culturally adapt the Lower Extremity Functional Scale in modern standard Arabic that is widely used and understood in the Middle East and North Africa region, and assess its psychometric properties.

Methods

Cross-cultural adaptation followed a combination of recommended guidelines. For psychometric evaluation 150 patients with anterior cruciate ligament injury and 65 asymptomatic individuals were recruited. All measurement properties as indicated by the Consensus-based Standards for the selection of health status Measurement Instruments recommendations were evaluated, including content-relevance analysis, structural validity, longitudinal reproducibility, anchor- and distribution-based methods of responsiveness, as well as the longitudinal pattern of change of Lower Extremity Functional Scale in anterior cruciate ligament injured patients' functional status.

Results

The questionnaire presented excellent internal consistency ($\alpha = 0.96$), reliability (0.80–0.98), and good convergent validity ($\rho = 0.85$). For reproducibility testing: minimal detectable change was 9.26 points; for responsiveness assessment: minimal clinically important difference was 9 points and presented moderate effect sizes (Glass' $\Delta = 0.71$, Cohen's $d = 0.81$). Its unidimensionality was not confirmed and an exploratory factor analysis indicated a 2-factor solution explaining 78.1% of the variance.

Conclusion

The Arabic Lower Extremity Functional Scale presented acceptable psychometric properties comparable to the original version. The Arabic version of Lower Extremity Functional Scale can be used in research and clinical practice to assess the functional status of Arabic-patients suffering an anterior cruciate ligament injury.

Introduction

The Lower Extremity Functional Scale (LEFS) is a patient-reported outcome measure which assesses the functional status of patients who have lower extremity conditions of musculoskeletal origin. The 20-item questionnaire inquires as to the degree of difficulty of performance of each of these activities on a 5-point Likert scale and a score of 80 represents an individual with normal function. [1]

Despite that the number of available patient-reported outcome measures has increased dramatically over the past decades, most of these instruments are developed for English-speaking patients. These tools in order to be used in different language and culture populations require a specific methodology with aim the adequate linguistic translation, but more importantly the cultural adaptation to maintain the content validity of the instrument across different cultures. [2] The LEFS utilizes lay terminology in simple English sentence format making it feasible to translate into modern standard Arabic for a lay population. Regional Arabic dialects often create barriers to clear communication whereas modern standard Arabic is widely used and understood in the Middle East and North Africa region (MENA). Recently, an Arabic version of LEFS [3] was developed, but it was only tested in Saudi Arabian participants. The authors suggested formal testing of the measurement properties of the tool before administration in patients from different Arabic countries. [3] We tested the questionnaire in our institution and we noticed that the translation was literally acceptable but confusing when applied to different Arabic-speaking ethnicities or to patients with low educational level. Specifically, we observed that the researchers maintained culturally irrelevant items and terminology not usually used in Arabic countries or not easily comprehensible by lay population. Accordingly, we decided to develop a version of the questionnaire maintaining its content validity and being understood across most of the Arabic speaking world.

LEFS has been validated in several musculoskeletal conditions and its context-dependent measurement properties have been recently evaluated and summarized. [4] However, the LEFS has been used infrequently in studies examining patients following Anterior Cruciate Ligament (ACL) reconstruction and its psychometric properties have not been determined in this population. [4, 5] Therefore, the main objectives of this study were: i) to translate and

cross-culturally adapt the LEFS for a wide spectrum of Arabic-speaking patients and ii) to evaluate its psychometric properties in a cohort with an ACL injury.

Methods

The translation and cross-cultural adaptation process adhered to published guidelines. [2, 6, 7] The validation of LEFS followed quality criteria on the evaluation of health status questionnaires [8] and the Consensus-based Standards for the selection of health status Measurement Instruments (COSMIN) recommendations. [9]

This study was conducted at the rehabilitation department of our institution. Ethics approval was obtained from the Institutional Review Board (Anti-Doping Lab Qatar—SCH-ADL-A-071), all participants gave written informed consent, and the study was conducted from 2013 to 2016.

Translation and cross-cultural adaptation

The original LEFS [1] developed in English language and was translated into modern standard Arabic. To ensure uniformity between source and target version special attention was paid to semantic, idiomatic, experiential, and conceptual equivalence. [2] The process followed 8 steps merged from published recommendations [2, 6, 7] (Table 1). Subsequently, proposed psychometric properties [8, 9] to ensure quality of the evaluation of the Arabic LEFS (LEFS-MSAr) were established (S1 File).

Sample size calculation and participants

The sample size required for the study was based on the intraclass correlation coefficient (ICC) and the maximum width of the 95% confidence intervals among studies assessing the psychometric properties of the scale. [4] The formula used to calculate the sample size [10] was $n = 16p(1-p)/w^2$, where p was the lowest expected ICC (0.85) and w was the maximum reported width (0.15) of the 95% confidence interval. [4] The sample size was calculated to be 91; however 215 individuals were prospectively recruited. We enrolled 150 male ACL patients (27.1 ± 7.0 years) for the following reasons: i) a minimum number of 100 subjects is required to ensure stability of the variance-covariance matrix in dimensionality analysis, [8] and ii) accounting for non-attendances at rehabilitation sessions to ensure that reproducibility testing would be done in “stable” patients. [8] Moreover, we enrolled 20 healthy (19.4 ± 1.5 years) and 45 “at risk” (soccer players) for an ACL injury individuals (22.7 ± 3.6 years) to evaluate interpretability of the scale. [8] These participants were recruited through direct contact during their training sessions at their sport clubs (during recruitment 3 potential candidates for the healthy and 8 potential candidates for the “at risk” group declined to participate).

Inclusion and exclusion criteria

All participants had to be at least 18 years of age and willing to provide written informed consent. Patients were eligible for the study if they had an ACL injury and either had a reconstruction or were undergoing conservative care. Participants were excluded if they had other orthopaedic conditions (e.g. low back pain, referred spinal symptoms, immobilised fracture) that would significantly compromise their physical functional status. For the healthy and “at-risk” groups, additional exclusion criteria were pain and functional deficits of the lower limbs during sports participation. All asymptomatic individuals who met these criteria and assessed by a physiotherapist were included.

Table 1. The process of translation and cross-cultural adaptation of the LEFS questionnaire for Arabic-speaking patients.

Steps	Procedures
Step 1: <i>Initial Translation</i>	Two bilingual and bicultural translators, whose native language was Arabic, independently produced 2 translations and 2 written reports. One translator (informed) had medical background (physiotherapist) and was aware of the construct of the scale, while the other translator (naïve) had no clinical background (language teacher) but was knowledgeable about the cultural and linguistic nuances of the Arabic language.
Step 2: <i>Reconciliation Committee</i>	The pair of translators in collaboration with a bilingual committee (3 physiotherapists and a sports medicine physician), a coordinator (researcher with several years of experience in scales development and validation), and a recording bilingual researcher synthesized the 2 translations and through a consensus process harmonized and produced a common initial translation and a written report documenting the synthesis process.
Step 3: <i>Back Translation</i>	Two bicultural translators, whose native language was English and who were fluent in the target language, produced 2 independent back translations of the initial questionnaire. Both were uninformed of the concepts explored to avoid information bias, had no medical background, and were blind to the original questionnaire.
Step 4: <i>Back translation review committee and harmonization</i>	An independent committee with aim the conceptual equivalence of translation consisting of the translators, members of the research team, and included bilingual clinicians knowledgeable about the content area convened, reached consensus, and developed the pre-final version of the LEFS-MSAr for translation validation. During this process the committee assessed the original questionnaire [1] and each translation (reconciled and back) together with the corresponding written report. Special attention was given by the committee to ensure intertranslation validity and to achieve semantic, idiomatic, experiential, and conceptual equivalence between the source and target questionnaire. [2, 6, 7] Furthermore, the committee discussed “comparability of language” which refers to the formal similarity of words and sentences between the original and back-translated questionnaire, as well as “similarity of interpretability” which refers to the degree to which the versions produce the same response even if the wording differs. [2, 6, 7]
Step 5: <i>Validation of translation</i>	The validation of the LEFS-MSAr regarding the success of the translation process was assessed in two ways: a) Formal evaluation of comparability of language and similarity of interpretability by using 7-point Likert scales ranging from 1 (extremely comparable/extremely similar) to 7 (not at all comparable/not at all similar). [6] Six bilingual individuals (3 men and 3 women, mean age 30.8 years and range 26 to 35 years) rated each original and back-translated item. Following this process each mean score >3 (comparability) and >2.5 (similarity) requires review for possible correction. b) Cognitive debriefing: the LEFS-MSAr was tested for cognitive equivalence [7] by 10 native Arabic-speaking patients at various educational levels representing the Gulf region population (Qatar, UAE, Jordan, Syria, Lebanon, Morocco, Tunisia, and Egypt) in order to capture the differences in Arabic dialects across the Middle East/North Africa region (mean age _(range) 30.6 _(19–45) years). The same 10 individuals were also used for the item-content relevance analysis (see content validity).
Step 6: <i>Review and finalization</i>	Considering the comments from the former process the committee made all necessary modifications for improvement.
Step 7: <i>Proofreading</i>	A proofreading company checked the final version for spelling, diacritical, grammatical, or other errors.
Step 8: <i>Pretesting</i>	The pre-final version of the LEFS-MSAr was administered to 20 Arabic-speaking patients suffering from lower limb musculoskeletal conditions (14 men and 6 women, with age _(range) of 36.6 _(19–59) years). Following the completion of the questionnaire, each individual was formally interviewed regarding the comprehension of items and the chosen response as part of the assessment of face and content validity. Upon completion of pre-testing a committee convened and the pre-final version without corrections was accepted as the final version of the LEFS-MSAr questionnaire (S1 File).

<https://doi.org/10.1371/journal.pone.0217791.t001>

Procedures, measurement and psychometric properties

The LEFS-MSAr was administered to all participants (n = 215) and completed twice within 3.3 ±2.0 days in the presence of one of the investigators. On completion of the questionnaire if the investigator identified a missing item the patient was questioned why. If it was an oversight, the patient was asked to respond to the item. Where the item was as not-applicable an explanation was requested and recorded. This interval between test-retest was chosen as the time period short enough to guarantee that clinical change has not occurred. [5, 8] Furthermore, based on previously published methodology, [11, 12] to ensure stability of the condition we only included participants who self-rated their condition as unchanged at the second assessment occasion.

The assessed measurement properties of the LEFS-MSAr questionnaire are presented in Table 2.

Table 2. Measurement and psychometric properties assessed for the LEFS-MSAr questionnaire.

Validity Testing	
Face validity [13–15]	Face validity of LEFS-MSAr was assessed: a) in 3 steps of the translation/ cross-cultural adaptation process (validation of translation, review and finalization, and pretesting), b) during the content analysis procedure (see content validity), c) by the participants that appraised the extent to which the instrument assessed their condition after completion of the questionnaire, and d) by the authors.
Content validity [13, 15]	Content validity was tested in all 20 items of LEFS-MSAr through a structured content analytic method. [16] The items were distributed to 10 judges (native Arabic-speaking patients) during the fifth step of translation and another 8 judges (1 academic, 2 sport physicians, 2 physiotherapists, and 3 professional athletes; all holding higher degrees in relevant areas) after the pretesting process. The judges matched each of the 20 items based on their content to a five-point Likert scale (1 = poor, 2 = fair, 3 = good, 4 = very good, or 5 = excellent match). This quantified item ratings allowing evaluation of judgments using quantitative statistical procedures. Finally, the content validity was also assessed by the 20 patients participated in the pre-testing phase of cross-cultural adaptation.
Construct validity [13–15]	Construct validity was evaluated by convergent validity. Concerning LEFS, a criterion scale does not exist; instead, the International Knee Documentation Committee subjective knee form (IKDC) [17] was the only region-related scale available for Arabic-speaking patients (translated and adapted, not published) for which we expected a high correlation (>0.80) with the LEFS based on a previous cross-cultural adaptation [18] that enrolled patients with a variety of knee disorders. Patients with ACL injury (n = 50, 28.6±8.4 years) during the initial assessment and before the administration of LEFS-MSAr were asked to complete the IKDC questionnaire.
Structural validity [13, 15]	Following guidelines [8] and given that the unidimensionality of LEFS has been confirmed, [1, 4] the structural validity as part of construct validity of LEFS was examined by using firstly confirmatory factor analysis (CFA). A poor model fit would be followed by an exploratory factor analysis (EFA) in order to establish its dimensionality and whether the items in the questionnaire group together in a consistent and scrutable way.
Known groups validity [13, 15]	Hypothesis-testing as part of construct validity evaluation was conducted by known groups validity and using two approaches: a) the contrasted-groups approach was tested by sampling the LEFS-MSAr mean scores of three distinct groups (early, intermediate and advanced stage of ACL rehabilitation) known to be high, mid, and low in the construct being measured [5] (n = 150), and b) by comparing the total LEFS score of the participants' groups. The healthy/"at risk", and patient group are distinct subgroups that are estimated to score differently in the LEFS (n = 215). We expected a higher score at healthy and "at risk" groups compared to the patient group.
Reliability Testing	
Inter-item reliability [13, 15]	The inter-item reliability of the LEFS was assessed by using Cronbach's alpha coefficient (α) that is considered to be the optimal estimate of the internal consistency and structural validity of the instrument where the scale is unidimensional. [15]
Test-retest reliability [8, 13–15]	To evaluate test-retest reliability the scale was administered to all the participants (n = 215) twice within 3.3±2.0 days. The time interval between repeated administrations should be long enough that the respondents do not recall their original responses, but short enough to ensure clinical stability of the condition. We re-tested only patients that self-rated their condition as unchanged between administrations.

(Continued)

Table 2. (Continued)

Validity Testing	
Longitudinal reproducibility [19, 20]	A measuring instrument should be reproducible while assessing changes in a given condition (longitudinal reproducibility). [19] In the present study we repeated the test-retest procedure (range 3 to 5 days) in a group of ACL patients (n = 20) after 8 weeks from initial administration of the LEFS-MSAr.
Utility evaluation	
Feasibility and acceptability [13, 14]	To appraise the acceptability and the ease of administration of the LEFS-MSAr we recorded the time spent by the participants filling it out and the percentage of unanswered questions.
Other properties	
Responsiveness [8, 20, 21]	Responsiveness reflects the ability of a questionnaire to detect clinically important changes over time or the validity of the change score. The receiver-operating-characteristic (ROC) curve analysis was used to determine the LEFS-MSAr change scores that best discriminate between the patients who were reported to achieve their goals and change group of rehabilitation (anchor) and the patients who did not change group.
Interpretability [8, 15]	Interpretability is defined as the degree to which one can assign qualitative meaning to an instrument's quantitative scores or change in scores. [8, 15] To aid in interpreting scores on LEFS-MSAr: a) normative scores of healthy and "at risk" for an ACL injury individuals were presented, b) the population of post-operative and conservatively managed patients was divided into three groups (early, intermediate, and advanced) by the treating physiotherapists using strict functional objective criteria (S1 Table); individual scores were presented and compared to their group/functional status, and c) we compared the scores of patients between two time points following treatment of known efficacy. It was expected based on published data [1, 5] that the LEFS scores would increase with days since surgery or injury, since one can reasonably expect patients to recover over time. We expected at least 9-point increase for a true difference according to previously published data from ACL patients. [5] More specifically, distribution-based methods were used to examine the magnitude of the change in the LEFS-MSAr and not the validity of the change score. The LEFS-MSAr was administered to a group of ACL patients (n = 90; 26±6.2 years) on two occasions 2 months apart. We hypothesized that clinically meaningful differences should be displayed in scores (higher scores) obtained by these patients as a result of rehabilitation. Minimal clinically important difference should not be considered a fixed property of the instrument; but we added this property in our analyses in order to interpret the change of scores over time in that specific population. [8, 20] Several strategies have been proposed for relating instruments' scores to independent standards that are themselves interpretable and serve as anchors, such as patient-rated global change or functional measures. [20] We calculated the Guyatt responsiveness index [19], as well as we implemented four anchor-based methods: a) the comparison of change scores with the Minimal detectable change (MDC ₉₅) as was calculated from the LEFS-MSAr scores that represents the true change beyond the error of the measurement, b) we compared the change in total scores with the smallest real difference (SRD) as was calculated from the mean change score of the 'stable' patients at 2-month follow-up (patients that did not change rehabilitation group). [22]
Ceiling and floor effects [8]	The LEFS-MSAr would be considered to have ceiling and floor effects if more than 15% of the patients scored the maximum and minimum possible score respectively. Relative to each item of the questionnaire ceiling and floor effects were considered to have occurred if at least 75% of the patients scored the maximum or minimum score to that item, respectively.

<https://doi.org/10.1371/journal.pone.0217791.t002>

Statistical analyses

Statistical analyses were performed using SPSS v19.0 and AMOS v24.0. The level of significance was set at $p < 0.05$. Descriptive statistics were used to calculate the characteristics of the participants, the scores of LEFS-MSAr and International Knee Documentation Committee subjective knee form (IKDC) questionnaires, the mean of days between test-retest, the mean scores for each item for comparability of language and similarity of interpretability, acceptability, and the ceiling and floor effects. Missing values were listwise excluded.

Validity testing. For item-content relevant analyses the judges' ratings were evaluated based on the validation procedure of Aiken's item-content validity coefficient (V). [23] The V statistic provides statistical significance of judges' ratings about an item's content-match with its construct and its values range from 0 to 1 (1 = perfect agreement). The values were then

compared against a right-tailed binomial probability table provided by Aiken [23] (V scores >0.70 considered as having acceptable validity, $p < 0.01$). Convergent validity was assessed with Spearman ρ (r) between the scores obtained from LEFS-MSAr and IKDC subjective knee form. [17]

A confirmatory factor analysis (CFA) on the LEFS items was conducted to evaluate the fit of the data to the hypothesised one-factor model. [1] The model parameters were estimated using the maximum likelihood method. The model was assessed by the standardized root mean square residual (SRMR), the root mean square error of approximation (RMSEA) (values <0.05 = good, <0.08 = adequate, >0.08 = poor fit for both indices), the comparative fit index (CFI) and the Tucker Lewis index (TLI) (values >0.95 = good, >0.90 = adequate, <0.90 = poor fit for both indices). [24] To explore the factorial validity of LEFS an exploratory factor analysis (EFA) (principal axis factoring) with varimax rotation was used. Eigenvalues over 1 were chosen and extracted, and items loading more than 0.40 were regarded as loading on a specific factor. Items loading more than 0.40 on 2 factors were assigned to the factor with a higher correlation. [25]

Known groups validity and group differences were calculated using the Kruskal-Wallis test, with post hoc comparisons using the Mann-Whitney U -test with Bonferroni correction for multiple testing, resulting from the formula $k(k-1)/2$, where k is the number of groups ($p_{adj} = 0.017$).

Reliability testing. The interrelatedness among the items of the LEFS was assessed by using Cronbach's α . Values of 0.70–0.90 have been proposed as a measure of good internal consistency. [8] Reproducibility was evaluated by using both Spearman's ρ and 2-way random effects model Intraclass Correlation Coefficient, type agreement ($ICC_{2,1}$), because systematic differences are considered to be part of the measurement error. [8, 26] As a measure of agreement the absolute measurement error was expressed as the standard error of measurement ($SEM_{Agreement} = SD \times \sqrt{1-ICC}$), including the systematic differences in order to distinguish them from real changes, e.g., due to treatment. [8] In addition, the minimal detectable change ($MDC_{95} = 1.96 \times \sqrt{2} \times SEM$) was calculated, which corresponds to the minimal within-person change in score that, with $p < 0.05$, can be translated as a real change above measurement error. [8, 27] Bland-Altman methods were used to indicate absolute agreement for test-retest measurements including a scatter plot of differences between applications, with 95% limits of agreement (mean change in scores of repeated administrations). [28]

Responsiveness and interpretability. The Wilcoxon test, using scores separated by 2 months was conducted to add in instrument's interpretability. Also, effect size (ES) by using both baseline and pooled standard deviation (SD) and standardised response mean (SRM) were calculated [20] and interpreted according to published recommendations (values of 0.20, 0.50, and 0.80 or greater represent small, moderate and large responsiveness, respectively). [29]

The Guyatt responsiveness index (GRI) was calculated as the ratio of the average change in improved patients divided by the SD of the change in stable patients as indicated by the rehabilitation stage. If the GRI was larger than 1, we considered the magnitude of change as acceptable. [19] Additionally, we estimated the magnitude of change (minimal clinically important difference—MCID) that is considered important to distinguish groups and to evaluate change over time by comparison of change scores with MDC_{95} , smallest real difference (SRD), The receiver-operating-characteristic curve (ROC), and area under the curve (AUC). Accordingly, to add in the interpretability of LEFS-MSAr we calculated the proportion of patients with a change score equal to or larger than the MDC_{95} for improved patients, or based on the SRD calculated ($SRD = 1.96 * SD_{change}$) [8] from the patients that were stable at the follow-up (according to rehabilitation groups' allocation—S1 Table).

The ROC curve was also used to provide the true-positive rate (sensitivity) versus the false-positive rate (1 – specificity). The most upper left point in the diagram represents the optimal cut-off change score, which most effectively discriminates between patients who have improved and those whose condition is unchanged. [22, 30] Additionally, the AUC reflects the probability of correctly discriminating between improved and non-improved patients. This area varies from 0.5 (the questionnaire does not discriminate more effectively than chance) to 1.0 (perfect discrimination). [22, 31]

Results

Translation and cross-cultural adaptation

During the steps of harmonization and validation of translation the native Arabic members of the committee and individuals used for formal evaluation and cognitive debriefing revealed a potential for 5 activities that were not entirely relevant to the Arabic culture and required modification. At this point, in order to achieve cultural content equivalence, experts in physiotherapy that have an intimate knowledge of the cultural habits across the Middle East North Africa region were also consulted. Discussions were held, and a consensus formulated that 5 items would be changed (Table 3). Following these changes, the steps 4, 5 and 6 of the translation process (Table 1) were repeatedly conducted as required. These changes were distributed to the original authors of the LEFS [1] as well as other authors who have published validation studies of the North American version of the LEFS, [32] and the changes were found to be acceptable (Alcock K, personal communication 2013).

One change was required in the instructions for the patient using the LEFS to clarify the meaning in Arabic of “do you or would you have any difficulty”. Consensus was reached among the committee and a small change was made to the instruction. The other two translation issues were not as easily resolved. Item “c” asks about Arabic sitting on the floor and item “f” asks about squatting. Both activities do not have a sufficiently detailed single word or phrase in Arabic that is consistent across all Arabic speaking countries or in modern standard Arabic. Additionally, when the tool was presented for validation of translation, these 2 questions were consistently highlighted as unclear in what was being asked. Consequently, the decision was made to partially abandon a literary description in favour of an image that clearly demonstrates the activity referred to in the question. Distribution of the LEFS-MSAr with the images was unanimously preferred among all the participants from the target population with no other changes required.

Validity testing

An overview of the measurement properties of both the LEFS-MSAr and the original version are presented in Table 4.

The face validity of the questionnaire was rated as excellent from participants at pre-testing, expert committees, judges at item content analyses, ACL injury patients, and authors. Content

Table 3. Items of the LEFS modified to suit for Arabic culture.

Item	Original Item (from North American Culture)	Modified item suitable for Arab Culture
c	Getting into or out of the bath	Arabic sitting on the floor
e	Putting on your shoes and socks	Kneeling on the floor to pray
k	Walking 2 blocks	Walking 200 meters
l	Walking a mile	Walking 1.5 kilometers
o	Sitting for 1 hour	Sitting in a chair for 1 hour

<https://doi.org/10.1371/journal.pone.0217791.t003>

Table 4. Summary of measurement properties of the original LEFS and LEFS-SMAr questionnaire.

Measurement property	LEFS-SMAr	LEFS-original*
Convergent validity†		
ACL patients–LEFS first assessment (n = 50)	$\rho = 0.85, p < 0.001$	r = 0.80
ACL patients–LEFS second assessment (n = 50)	$\rho = 0.88, p < 0.001$	
Factorial validity		
Factor structure	2-factor solution	1-factor solution
Variance explained	78.145%	
Communalities	0.695–0.906	
Factor loadings	0.58–0.85	0.58–0.86
Internal consistency		
Cronbach's α (n = 150)	0.965	0.96
Cronbach's α ADL	0.982	
Cronbach's α strenuous activities	0.934	
Test-retest reliability		
ICC (n = 215)	0.983 (95%CI: 0.977–0.987)	0.94 (95%CI lower 0.89)
ICC ACL (n = 150)	0.976 (95%CI: 0.964–0.983)	
ICC “at risk” (n = 45)	0.869 (95%CI: 0.772–0.926)	
ICC healthy (n = 20)	0.803 (95%CI: 0.685–0.921)	
Test-retest reliability		
Spearman's rho ACL	$\rho = 0.978, p < 0.001$	Not reported
Spearman's rho “at risk”	$\rho = 0.840, p < 0.001$	
Spearman's rho healthy	$\rho = 0.918, p < 0.001$	
Longitudinal reproducibility		
ICC ACL (n = 20)	0.988 (95%CI: 0.969–0.995)	Not reported
Agreement		
SEM ACL	3.34	3.9
SEM “at risk”	0.64	
SEM healthy	1.25	
Minimal Detectable Change		
MDC95(%) ACL	9.26 (11.57)	9.0 (MDC90)
MDC95(%) “at risk”	1.77 (2.21)	
MDC95(%) healthy	3.46 (4.32)	
Interpretability		
ES using baseline SD	0.71 (Glass' Δ)	Not reported
ES using pooled SD	0.81 (Cohen's d)	Not reported
SRM	1.17	Not reported
GRI	2.2	Not reported
SRD	16.1	Not reported
MCID	9 points	Not reported
Responsiveness		
AUC (95%CI)	0.781 (0.676 to 0.886)	Not reported

*The original LEFS was tested in general lower extremity musculoskeletal conditions.

†Convergent validity was assessed using IKDC as comparator in the Arabic version, while in the original version the SF-36 was used.

Abbreviations: LEFS, lower extremity functional scale; ACL, anterior cruciate ligament injury; ADL, Activities of daily living; ICC, intraclass correlation coefficient; SEM, standard error of measurement; ES, effect size; SD, standard deviation; SRM, standardized response mean; GRI, Guyatt responsiveness index; SRD, smallest real difference; AUC, area under the curve; MCID, minimal clinically important difference.

<https://doi.org/10.1371/journal.pone.0217791.t004>

validity was assessed through a structured content analytic method [16] and characterised as being well addressed by all 10 judges and 20 patients at pre-testing. All 20 items presented V values ranging from 0.725 to 1.0 ($p < 0.01$). Convergent validity was demonstrated by a high association between the scores of LEFS-MSAr and IKDC Subjective Knee Form [17] as expected (both test and retest $\rho > 0.80$).

Structural validity. The data from 215 participants were used in the CFA. The one factor model solution exhibited poor fit ($\chi^2 = 1563.77$; $\chi^2/df = 9.2$; SRMR = 0.0878; RMSEA = 0.195 (0.186–0.204), $p_{close} < 0.001$; CFI = 0.744; TLI = 0.714). CFA suggested that more than one factor underlie the LEFS-MSAr items in an ACL injury population (Fig 1). In the EFA analysis, the Kaiser-Meyer-Olkin measure (KMO) verified the sampling adequacy for the analysis (KMO = 0.956), which is well above the acceptable limit of 0.5. [33] Bartlett's sphericity test ($\chi^2(190) = 5435.096$, $p < 0.001$) indicated that the correlations between items were sufficiently large for EFA. The analysis and the examination of the screeplot (Fig 2) indicated a 2-factor solution with eigenvalues over Kaiser's criterion of 1 explained 78.14% of the total variance (Table 5). The items that cluster on the same components suggested that component 1 represents relatively light activities of daily living and component 2 sport related and strenuous activities for the knee. For that reason, we calculated internal consistency coefficient for each unidimensional subscale separately (Table 4).

Known group validity. Regarding normative values, Kruskal-Wallis tests revealed significant differences ($p < 0.001$) for mean LEFS-MSAr scores at first and second administration. No within group differences were found at both administrations. ACL injured patients scored significantly lower ($p < 0.017$) than both athletes "at risk" and healthy individuals (Table 6).

For contrasted-groups validity testing, Kruskal-Wallis test showed significant differences ($\chi^2(2) = 91.012$, $p < 0.001$) for the LEFS-MSAr scores with respect to the ACL rehabilitation stage, with a median score of 25.0 for "early", 53.0 for "intermediate", and 72.0 for "advanced" (Fig 3).

Reliability testing. Reliability results are presented in Table 4. The Cronbach's alpha for internal consistency was 0.965, and the Cronbach alpha if item deleted (for each item) varied from 0.962 to 0.965. The ICC_{2,1} total was 0.98 with the 95% confidence interval (CI) of 0.98 to 0.99. The SEM was calculated to be 3.34 points and the MDC₉₅ 9.26 points for the ACL injured patients. LEFS-MSAr was also proven longitudinally reproducible presenting an ICC_{2,1} of 0.99, a SEM of 1.42 points and a MDC₉₅ of 3.94 points.

A Bland-Altman plot (Fig 4) showed that the differences between two assessments were plotted around the zero line and within the limits of agreement (-5.5 to 7.7) with a few outliers. Moreover, the zero line was within the 95% CI of the mean difference indicating no systematic bias.

Utility evaluation

The completion of the questionnaire required 3 to 5 minutes maximum and none of the participants reported language comprehension or semantic problems. All participants completed the full LEFS-MSAr, resulting in the maximum response rate.

Other properties

Responsiveness. The ROC analysis indicated that LEFS-MSAr (Fig 5) moderately discriminated improved and non-improved patients with an AUC associated with this value of 0.781 (95%CI 0.676 to 0.886). The represented sensitivity and specificity were 0.85 of 0.65, respectively.

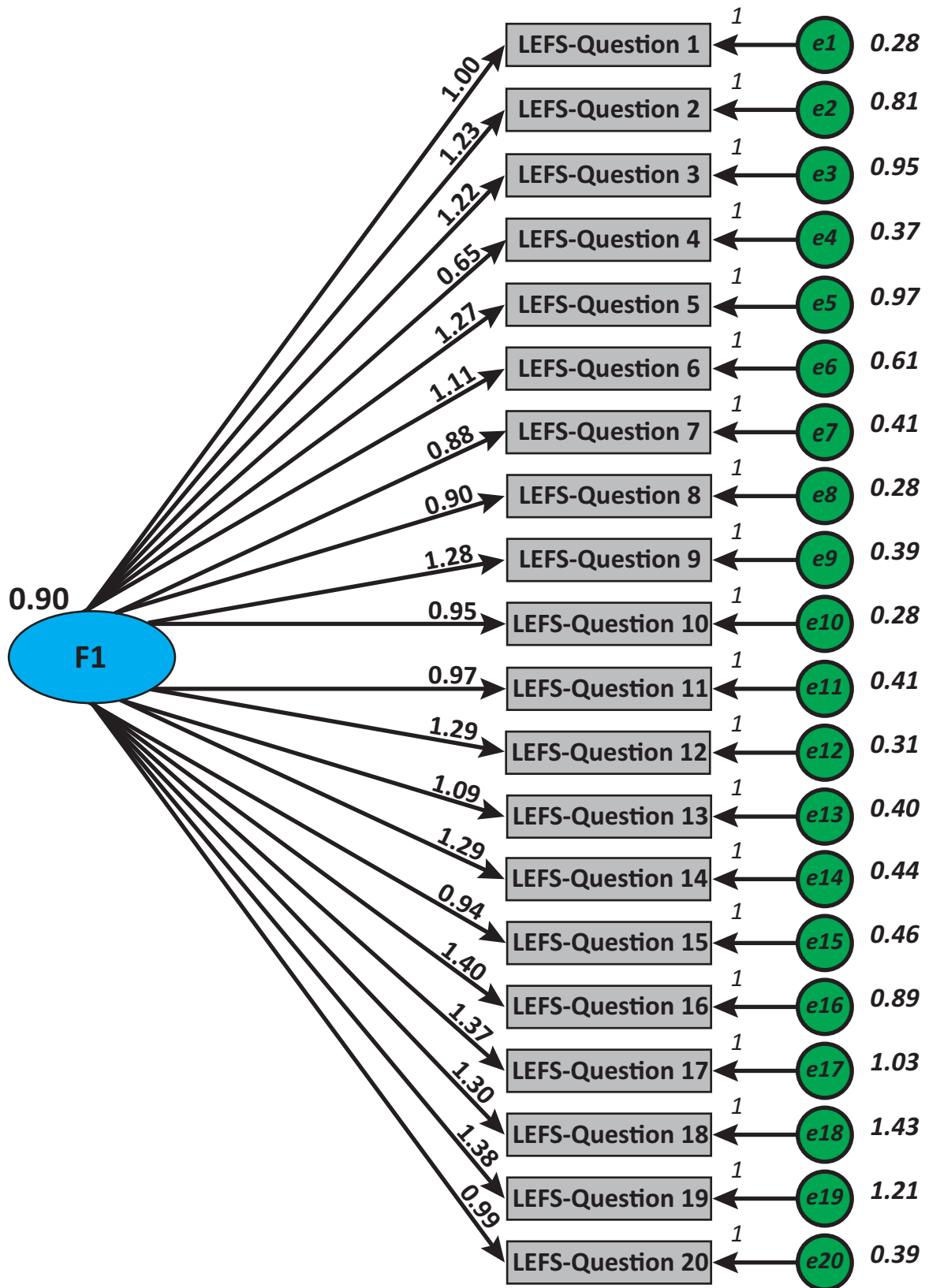


Fig 1. Confirmatory factor analysis. Model of the hypothesized 20-item 1-factor structure for the modern standard Arabic Lower Extremity Functional Scale (LEFS-MSAr).

<https://doi.org/10.1371/journal.pone.0217791.g001>

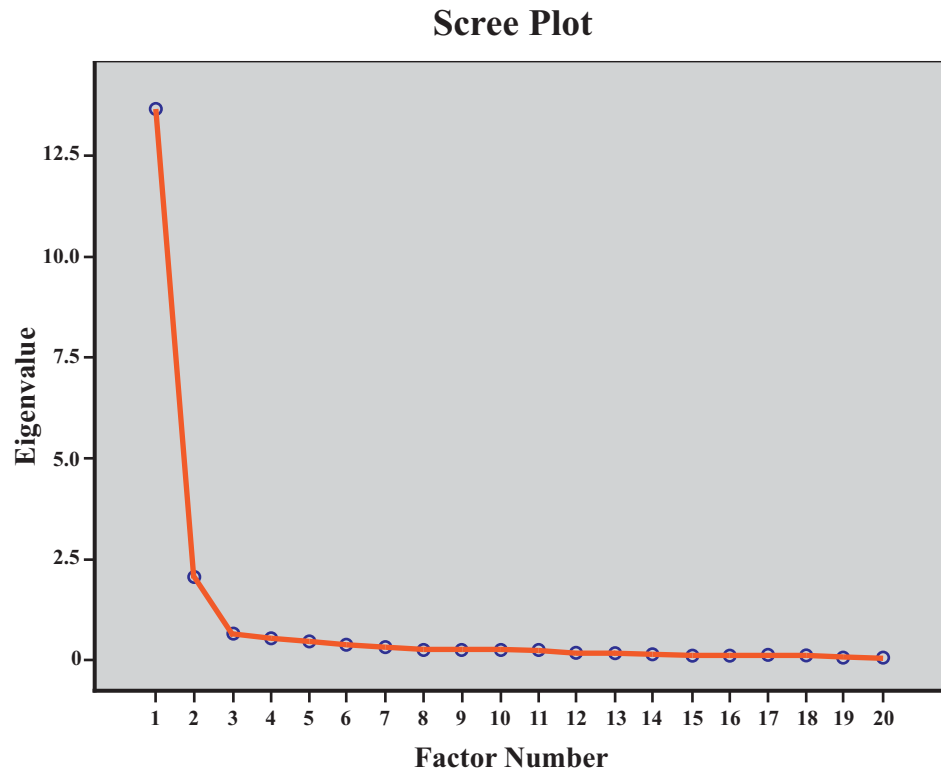


Fig 2. Exploratory factor analysis. Scree plot of eigenvalues form the 20-item modern standard Arabic version of Lower Extremity Functional Scale (LEFS-MSAr).

<https://doi.org/10.1371/journal.pone.0217791.g002>

Interpretability. The LEFS-MSAr was administered to a group of ACL patients ($n = 90$) on two occasions two months apart. Twelve patients were excluded due to subsequent ACL reconstructive surgery.

The Wilcoxon test revealed statistically significant changes of the LEFS-MSAr from first (Median = 50.5) to second (Median = 67.5) administration for this group ($Z = -7.62$, $p < 0.0001$) representing large effect sizes (Table 4).

Based on a change score equal or larger to the MDC_{95} at the re-administration of the LEFS-MSAr, 59.0% of the patients were rated as improved, while 41% were found with no change. However, at the final assessment 38 ACL patients were at the advanced (final) stage of rehabilitation where little further improvement is expected.

Based on change of rehabilitation group (early, intermediate, and advanced) 53.1% ($\text{mean}_{\text{change}} \pm \text{SD}_{\text{change}} = 18.1 \pm 12.6$) of the patients were rated as improved, while 46.9% ($\text{mean}_{\text{change}} \pm \text{SD}_{\text{change}} = 7.5 \pm 8.2$) were found with no change.

The SRD calculated from the change scores for the 'stable' patients was 16.1 points, indicating that an individual had to change at least 16 points on the LEFS-MSAr to be judged as having really changed. Based on SRD the sensitivity to change of the LEFS-MSAr was 36%. Seven of the stable patients (11%) had a change score higher than SRD, indicating a specificity to change of around 89%.

Normative scores of LEFS-MSAr are presented in Table 6, while individual patient scores according to functional status are presented in Fig 3. Significant changes over time were found for LEFS-MSAr ($p < 0.0001$) with a large ES following rehabilitation. Fig 6 documents the two-month test-retest data with patients classified according to their independently rated

Table 5. Exploratory factor analysis with varimax rotation suggesting a 2-factor solution.

Item	Rotated factor loadings	
	Activities of daily living	Sports and strenuous activities
h. Performing light activities around your home	0.85	0.31
k. Walking 200 meters	0.84	0.26
g. Lifting an object, like a bag of groceries from the floor	0.82	0.26
j. Getting into or out of a car	0.81	0.38
a. Any of your usual work, housework or school activities	0.80	0.39
t. Rolling over in bed	0.78	0.36
d. Walking between rooms	0.76	0.19
o. Sitting in a chair for 1 hour	0.74	0.34
m. Going up or down 10 stairs (about 1 flight of stairs)	0.74	0.43
k. Walking 1.5 kilometres	0.71	0.55
f. Squatting	0.68	0.44
i. Performing heavy activities around your home	0.67	0.57
n. Standing for 1 hour	0.67	0.55
r. Making sharp turns while running fast	0.20	0.92
q. Running on uneven ground	0.30	0.90
s. Hopping	0.29	0.86
p. Running on even ground	0.39	0.84
b. Your usual hobbies, recreational or sporting activities	0.42	0.76
e. Kneeling on the floor to pray	0.44	0.70
c. Arabic sitting on the floor	0.52	0.58
Eigenvalues	13.64	1.97
% of variance	68.27	9.87
Total Variance %	78.145	

<https://doi.org/10.1371/journal.pone.0217791.t005>

rehabilitation status. The mean change in LEFS-MSAr score was significant ($p < 0.01$) for those subjects who changed groups (22 and 14 points, respectively), and exceeded the documented MDC (9 points), whereas the change for those subjects who stayed in the same group was non-significant (8 points).

Ceiling and floor effects. No floor effect was found for LEFS-MSAr total score at first (0%) or second assessment (0%), nor a ceiling effect (1.3% and 2.6%, respectively).

Table 6. Total scores of the LEFS-SMAr questionnaire in the groups of participants.

Group	N	Test*	Re-test*
ACL	150	51.0 ± 21.2 (47.6–54.4)	52.4 ± 21.9 (48.9–55.9)
At Risk	45	78.6 ± 2.2 (78.0–79.3)	78.9 ± 2.0 (78.4–79.6)
Healthy active	20	78.3 ± 2.2 (77.3–79.3)	77.4 ± 3.4 (75.8–79.0)

*Data are presented as mean ± SD (95% CI)

The Mann-Whitney test did not reveal significant differences between healthy and at risk groups at both assessments ($U_{test} = 400.0$, $U_{retest} = 307.5$, $p > 0.017$ respectively), but ACL injury patients scored significantly lower at both assessments than at risk ($U_{test} = 244.0$, $U_{retest} = 307.5$, $p < 0.001$ respectively) and healthy group ($U_{test} = 130.5$, $U_{retest} = 264.5$, $p < 0.001$ respectively).

Abbreviations: LEFS-SMAr, lower extremity functional scale Arabic version; N, sample size; ACL, anterior cruciate ligament injury patients; healthy active, healthy active individuals.

<https://doi.org/10.1371/journal.pone.0217791.t006>

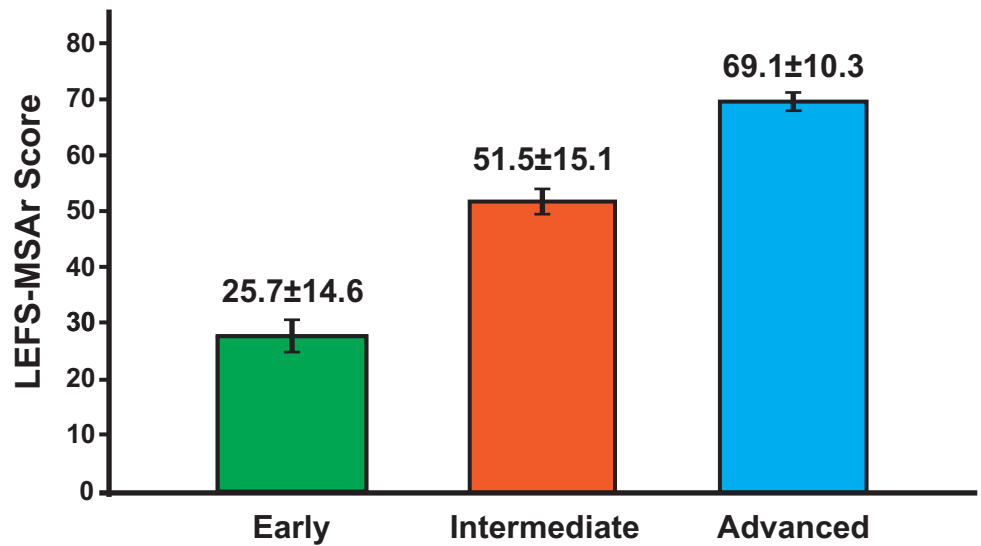


Fig 3. LEFS-MSAr scores according to the ACL rehabilitation stage. LEFS-MSAr mean(±SD) for each of the ACL in terms of rehabilitation stage groups. Mean total score and standard error values are depicted in the graph. Significant differences were found between scores for early (median = 25.0) and intermediate (median = 53.0) group ($U = 264.500, p < 0.001$), early and advanced (median = 72.0) group ($U = 24.000, p < 0.001$), and intermediate and advanced group ($U = 456.000, p < 0.001$). Abbreviations: LEFS-MSAr, lower extremity functional scale modern standard Arabic version; N, sample size; ACL, anterior cruciate ligament injury patients.

<https://doi.org/10.1371/journal.pone.0217791.g003>

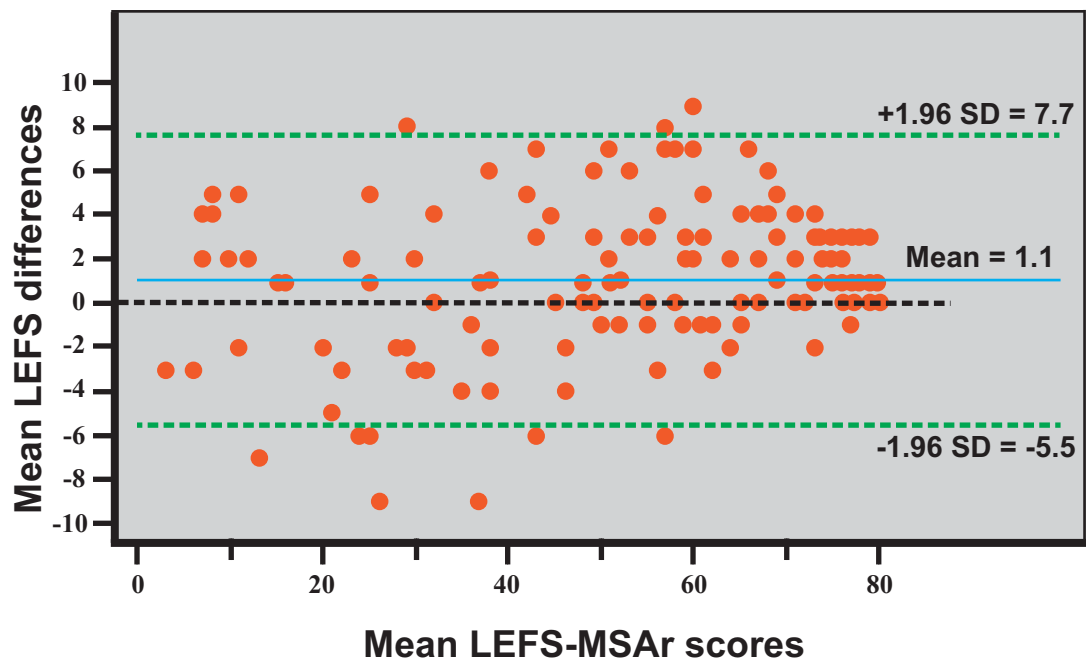


Fig 4. Bland-Altman plot. A Bland-Altman plot visualizing the agreement for test-retest, with the limits marked as mean±SD difference. Means and differences were calculated using total original scores of the scale. Abbreviations: LEFS-MSAr, lower extremity functional scale modern standard Arabic version.

<https://doi.org/10.1371/journal.pone.0217791.g004>

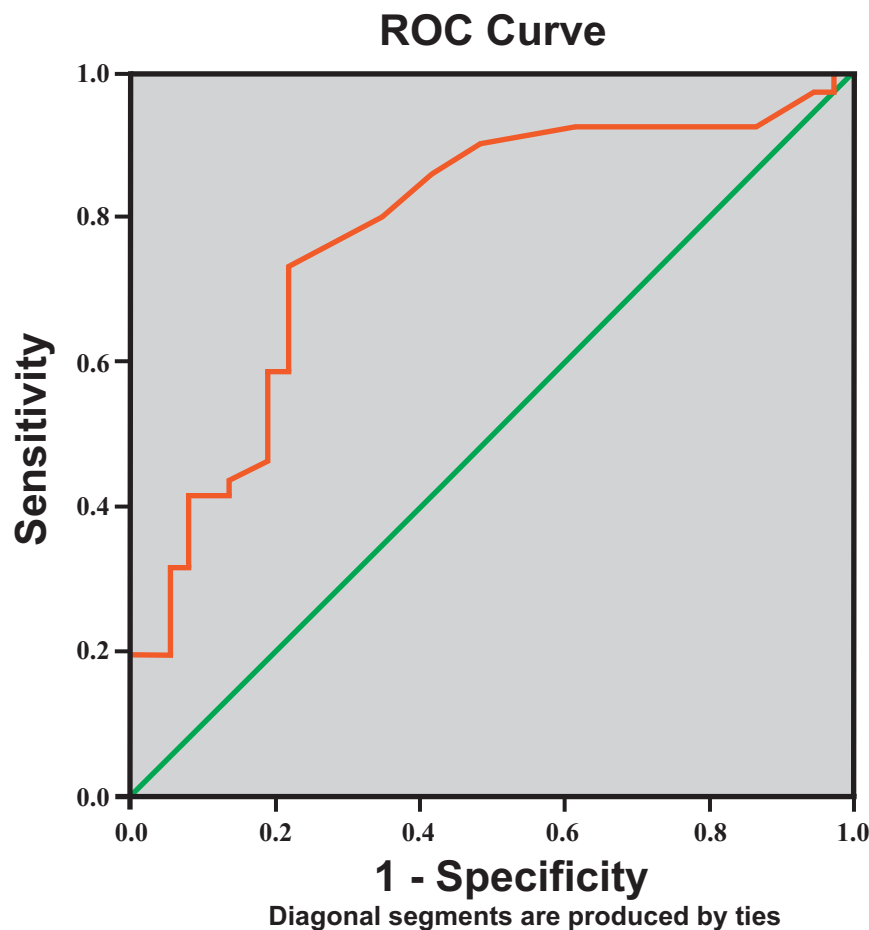


Fig 5. Receiver-operating-characteristic (ROC) curve. Receiver-operating-characteristic (ROC) curve illustrating the relationship between sensitivity and complement of specificity (1-specificity) for the modern standard Arabic version of Lower Extremity Functional Scale.

<https://doi.org/10.1371/journal.pone.0217791.g005>

Additionally, no individual items of the scale were scored at their maximum or minimum score by more than 75% of the patients at first (floor range 6–44.4%, ceiling range 6–64.2%) or second assessment (floor range 3.3–49%, ceiling range 13.9–74.2%).

Discussion

The LEFS was translated into modern standard Arabic, significant cultural adaptations were implemented, and the LEFS-MSAr demonstrated excellent psychometric properties as hypothesised in a cohort of patients with ACL injury.

Translation and cross-cultural adaptation

There are at least 12 major sets of guidelines available for questionnaires translation [7] and to our knowledge there is no consensus on a set of rigid procedures in the area of translation and cross-cultural adaptation. To overcome this problem and ensure translation quality we synthesised a rigorous process from published recommendations. [2, 6, 7]

A previous translation of the LEFS into Arabic [3], presented some difficulties in cultural relevance and comprehension. [6] Specifically, administering this version to our multinational

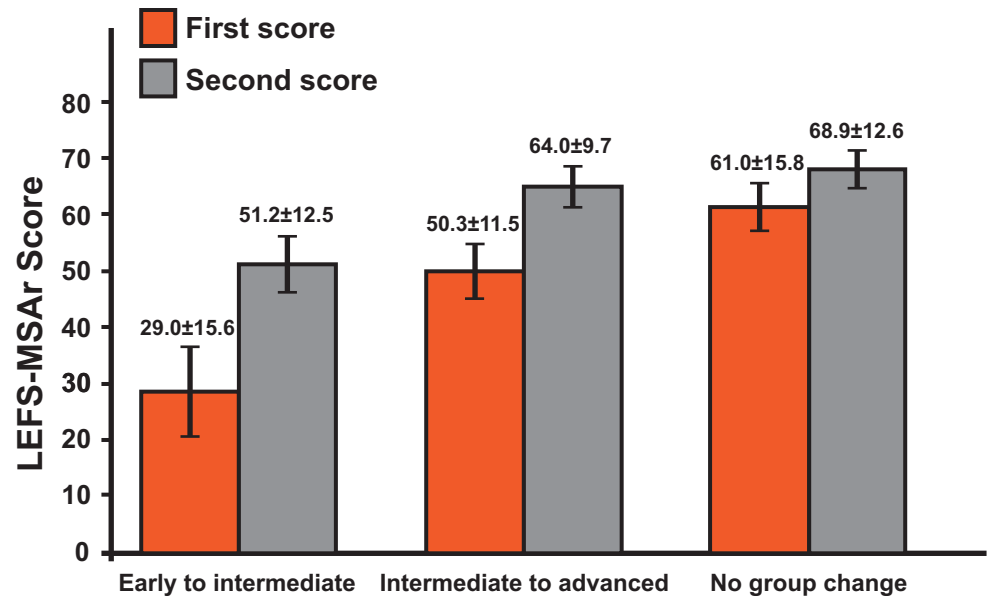


Fig 6. Mean LEFS-MSAr scores over 2 months. Mean LEFS-MSAr scores over a two-month test-retest, grouped by independently rated ACL rehabilitation group. Abbreviations: LEFS-MSAr, lower extremity functional scale modern standard Arabic version.

<https://doi.org/10.1371/journal.pone.0217791.g006>

Arabic-speaking patients we realized that patients were not able to understand some questions, as for example item “f” inquiring about squatting. Moreover, we observed that the item “c” was translated as “walking in and out of the bathroom” instead of “getting into or out of the bath”. Finally, some items were culturally irrelevant and had to be rephrased to reflect modern Arabic reality. Examples of problems the authors have encountered may be illustrative: i) item “o” inquiring about sitting does not reflect the Arabic culture of sitting (Arabic sitting on the ground), and ii) item “e” inquiring about “putting on shoes and socks” may not reflect an activity of older individuals that usually wear sandals. Accordingly, we cross-culturally adapted the LEFS into modern standard Arabic aiming to produce in a widely comprehensible and practical tool for use in Arabic culture across the Middle East North Africa region. Given that there are potential cultural differences in the interpretation of many terms, appropriate attention was given to cultural nuances by using bicultural members in the committees and implementing a formal translation validation by using native Arabic-speaking patients representing the Middle East North Africa region population.

Validity testing

As hypothesised, the LEFS-MSAr demonstrated good translational and construct validity. Evaluation of content validity by a structured analytic method [16] added to the psychometric properties of LEFS, as to our knowledge this is reported first time, [4] since most of previous studies used floor and ceiling effects to examine this form of validity. [4] The results also confirmed our hypothesis regarding the convergent validity of LEFS-MSAr with the IKDC Subjective Knee Form [17] scores presenting a high positive correlation as previously reported. [18]

Strength of the present study was the structural validity evaluation by initially using CFA followed by EFA. However, the CFA/EFA revealed a 2-factor structure that did not support the hypothesized unidimensionality as previously reported. [1, 3, 34–37] The unidimensionality of LEFS has been questioned by another study used CFA [38] and the factor complexity of

LEFS has been demonstrated before exhibiting 3 factors in patients with hip and knee osteoarthritis, [39] and 2 factors in patients with total hip and knee replacement [40] and general musculoskeletal disorders of the lower extremity. [41] The configuration of patient population and conditions involved may explain these discrepancies.

Reliability testing

An excellent test-retest reliability was demonstrated for ACL injured patients (ICC = 0.976) within the range (0.85–0.998) [4] of reported reproducibility in multiple conditions among studies. Also, we assessed the longitudinal reproducibility of LEFS-MSAr by repeating the test-retest procedure after 8 weeks from initial administration; a property that is almost completely neglected in the evaluation of instruments. [20] Again the LEFS-MSAr showed excellent temporal stability presenting an ICC of 0.988.

Regarding absolute reliability, the SEM calculated for our ACL injured patients scores was 3.34 points which is within the reported range in the literature (0.88–5.6), [4] and comparable to the SEM reported for ACL reconstructed patients (3.7) [5] and patients with various knee disorders (3.6). [18]

Interpretability and responsiveness

Clinically meaningful score differences with large effect sizes, an SRM of 1.17 and a GRI of 2.2 were displayed, reflecting ability of LEFS-MSAr to effectively distinguish changes over time. The ES reported herein (0.71-Glass' Δ and 0.81 Cohen's d) is far less than that in other studies [42–45] ranging from 1.26 to 3.33. [4] ES is strongly affected by the type of injury (i.e. acute or chronic) and the interval between test-retest and must be interpreted by clinicians with caution. A big enough interval between two evaluations, as for example in the study of Lin et al, [43] (6 months) and a sample configuration of patients with non-chronic conditions (>65%) combined with long evaluation interval as in Cruz-Diaz et al, [44] may lead in ES overestimation (3.33 and 2.3, respectively).

MCID should not be considered a fixed property of the instrument, nor responsiveness as a constant characteristic of a measure, [8, 20, 46] while different methods of MCID calculation can result in different values of MCID, particularly when baseline scores are characterized by different levels of severity. [47] In our ACL population LEFS-MSAr was found sensitive to change with an MCID of 9 points marginally not meeting the requirement of being greater than the MDC_{95} (9.26 points). Despite that, a change of greater than 9 points can make the clinician reasonably confident that this change is not due to error, but also is a clinically meaningful change. [1] In the majority of studies assessing psychometric properties of LEFS [1, 41, 48, 49] the MCID did not exceed the measurement error. Calculation of indices like MDC is dependent on the selected CI range with related z value and following the COSMIN recommendations [8] the MDC corresponding to 95% CIs is preferred. Taking this into account, not even in the original publication did the reported MCID exceed the MDC_{95} (10.8 points); these results indicate that further investigation into these properties in patients with lower limb conditions is necessary.

Limitations

The study sample is specific to ACL injured patients of male gender. The extent to which our results can be generalized to female or patients with other conditions is unknown and has to be elucidated in future studies. Also, our methodology considered only classical test theory; given the inconsistency in available literature regarding the unidimensionality of the scale a rigorous Rasch analysis is much needed to examine in detail the internal structure of the LEFS.

Given that the assumption of unidimensionality was not met, the reported Cronbach's alpha may have overestimated the internal consistency of the scale. Finally, a possible deficiency of the study is the usage of achievement of functional treatment goals in MCID evaluation compared to all previous studies that used global change rating scale.

Supporting information

S1 File. Modern standard Arabic version of LEFS (LEFS-MSAr).
(PDF)

S2 File. Anonymized data.
(XLSX)

S1 Table. ACL rehabilitation progression criteria.
(DOCX)

Author Contributions

Conceptualization: Vasileios Korakakis, Michael Saretsky.

Data curation: Michael Saretsky, Matthew C. Azzopardi, Jasenko Klauznicer, Abdallah Itani.

Formal analysis: Vasileios Korakakis, Rodney Whiteley, Giannis Giakas, Nikolaos Malliaropoulos.

Funding acquisition: Vasileios Korakakis.

Investigation: Vasileios Korakakis, Michael Saretsky, Matthew C. Azzopardi, Jasenko Klauznicer.

Methodology: Vasileios Korakakis, Rodney Whiteley, Omar Al Sayrafi, Giannis Giakas, Nikolaos Malliaropoulos.

Project administration: Vasileios Korakakis, Michael Saretsky, Abdallah Itani.

Resources: Vasileios Korakakis.

Software: Vasileios Korakakis.

Supervision: Vasileios Korakakis, Michael Saretsky, Rodney Whiteley, Matthew C. Azzopardi, Abdallah Itani.

Validation: Vasileios Korakakis.

Writing – original draft: Vasileios Korakakis.

Writing – review & editing: Vasileios Korakakis, Michael Saretsky, Rodney Whiteley, Matthew C. Azzopardi, Jasenko Klauznicer, Abdallah Itani, Omar Al Sayrafi, Giannis Giakas, Nikolaos Malliaropoulos.

References

1. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): Scale development, measurement properties, and clinical application. *Phys Ther*. 1999; 79(4):371–83. PMID: [10201543](https://pubmed.ncbi.nlm.nih.gov/10201543/)
2. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000; 25(24):3186–91. <https://doi.org/10.1097/00007632-200012150-00014> PMID: [11124735](https://pubmed.ncbi.nlm.nih.gov/11124735/)

3. Alnahdi AH, Alrashid GI, Alkhaldi HA, Aldali AZ. Cross-cultural adaptation, validity and reliability of the Arabic version of the Lower Extremity Functional Scale. *Disabil Rehabil.* 2016; 38(9):897–904. <https://doi.org/10.3109/09638288.2015.1066452> PMID: 26186622
4. Mehta SP, Fulton A, Quach C, Thistle M, Toledo C, Evans NA. Measurement Properties of the Lower Extremity Functional Scale: A Systematic Review. *J Orthop Sports Phys Ther.* 2016; 46(3):200–16. <https://doi.org/10.2519/jospt.2016.6165> PMID: 26813750
5. Alcock GK, Werstine MS, Robbins SM, Stratford PW. Longitudinal changes in the lower extremity functional scale after anterior cruciate ligament reconstructive surgery. *Clin J Sport Med.* 2012; 22(3):234–9. <https://doi.org/10.1097/JSM.0b013e31824cb53d> PMID: 22450593
6. Sperber AD. Translation and Validation of Study Instruments for Cross-Cultural Research. *Gastroenterol.* 2004; 126(1):S124–S8. <https://doi.org/10.1053/j.gastro.2003.10.016>
7. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, 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(2):94–104. <https://doi.org/10.1111/j.1524-4733.2005.04054.x> PMID: 15804318
8. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007; 60(1):34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012> PMID: 17161752
9. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Qual Life Res.* 2010; 19(4):539–49. <https://doi.org/10.1007/s11136-010-9606-8> PMID: 20169472
10. Stratford PWS G.F. Sample size estimation for the comparison of competing measures' reliability coefficients. *Physiother Can.* 2003; 55:225–9.
11. Korakakis V, Malliaropoulos N, Baliotis K, Papadopoulou S, Padhiar N, Nauck T, et al. Cross-cultural Adaptation and Validation of the Exercise-Induced Leg Pain Questionnaire for English- and Greek-Speaking Individuals. *J Orthop Sports Phys Ther.* 2015; 45(6):485–96. <https://doi.org/10.2519/jospt.2015.5428> PMID: 25927499
12. Malliaropoulos N, Korakakis V, Christodoulou D, Padhiar N, Pyne D, Giakas G, et al. Development and validation of a questionnaire (FASH—Functional Assessment Scale for Acute Hamstring Injuries): to measure the severity and impact of symptoms on function and sports ability in patients with acute hamstring injuries. *Br J Sports Med.* 2014; 48(22):1607–12. <https://doi.org/10.1136/bjsports-2014-094021> PMID: 25287515
13. Devon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh.* 2007; 39(2):155–64. <https://doi.org/10.1111/j.1547-5069.2007.00161.x> PMID: 17535316
14. Bannigan K, Watson R. Reliability and validity in a nutshell. *J Clin Nurs.* 2009; 18(23):3237–43. <https://doi.org/10.1111/j.1365-2702.2009.02939.x> PMID: 19930083
15. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010; 63(7):737–45. <https://doi.org/10.1016/j.jclinepi.2010.02.006> PMID: 20494804
16. Dunn JGH, Bouffard M, Rogers WT. Assessing item content-relevance in sport psychology scale-construction research: Issues and recommendations. *Meas Phys Educ Exerc Sci.* 1999; 3(1):15–36.
17. Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P, et al. Development and Validation of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med.* 2001; 29(5):600–13. <https://doi.org/10.1177/03635465010290051301> PMID: 11573919
18. Metsavaht L, Leporace G, Riberto M, Sposito MM, Del Castillo LN, Oliveira LP, et al. Translation and cross-cultural adaptation of the lower extremity functional scale into a Brazilian Portuguese version and validation on patients with knee injuries. *J Orthop Sports Phys Ther.* 2012; 42(11):932–9. <https://doi.org/10.2519/jospt.2012.4101> PMID: 23047028
19. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol.* 1992; 45(12):1341–5. [https://doi.org/10.1016/0895-4356\(92\)90194-R](https://doi.org/10.1016/0895-4356(92)90194-R) PMID: 1460470
20. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res.* 2003; 12(4):349–62. PMID: 12797708
21. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol.* 1989; 42(5):403–8. [https://doi.org/10.1016/0895-4356\(89\)90128-5](https://doi.org/10.1016/0895-4356(89)90128-5) PMID: 2659745

22. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. *Int J Technol Assess Health Care*. 2001; 17(4):479–87. PMID: [11758292](#)
23. Aiken LR. Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educ Psychol Meas*. 1985; 45(1):131–42. <https://doi.org/10.1177/0013164485451012>
24. Lt Hu, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999; 6(1):1–55. <https://doi.org/10.1080/10705519909540118>
25. Bortz J. *Statistik für Sozialwissenschaftler*. 4th ed: Springer Verlag; 1999.
26. McGraw KO, Wong SP. Forming Inferences about Some Intraclass Correlation Coefficients. *Psychol Methods*. 1996; 1(1):30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
27. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek ALM. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res*. 2001; 10(7):571–8. PMID: [11822790](#)
28. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1(8476):307–10. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8) PMID: [2868172](#)
29. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. New Jersey: Lawrence Erlbaum Associates; 1988.
30. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol*. 2002; 14(2):109–14. PMID: [11845014](#)
31. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiol*. 1982; 143(1):29–36. <https://doi.org/10.1148/radiology.143.1.7063747> PMID: [7063747](#)
32. Stratford PW, Hart DL, Binkley JM, Kennedy DM, Alcock GK, Hanna SE. Interpreting lower extremity functional status scores. *Physiother Can*. 2005; 57(2):154–62.
33. Field AP. *Discovering statistics using SPSS (and sex and drugs and rock 'n' roll)* 3rd ed. London: Sage; 2009.
34. Gabel CP, Melloh M, Burkett B, Michener LA. Lower Limb Functional Index: Development and Clinical Properties. *Phys Ther*. 2012; 92(1):98–110. <https://doi.org/10.2522/ptj.20100199> PMID: [22052947](#)
35. Hoozeboom TJ, de Bie RA, den Broeder AA, van den Ende CHM. The Dutch Lower Extremity Functional Scale was highly reliable, valid and responsive in individuals with hip/knee osteoarthritis: a validation study. *BMC Musculoskel Dis*. 2012; 13:117. <https://doi.org/10.1186/1471-2474-13-117> PMID: [22748143](#)
36. Hou W-H, Yeh T-S, Liang H-W. Reliability and validity of the Taiwan Chinese version of the Lower Extremity Functional Scale. *J Formos Med Assoc*. 2014; 113(5):313–20. <https://doi.org/10.1016/j.jfma.2012.07.032> PMID: [24746117](#)
37. Liang H-W, Hou W-H, Chang K-S. Application of the Modified Lower Extremity Functional Scale in Low Back Pain. *Spine*. 2013; 38(23):2043–8. <https://doi.org/10.1097/BRS.0b013e3182a826e8> PMID: [23963016](#)
38. Alnahdi AH. Confirmatory factor analysis of the Arabic version of the Lower Extremity Functional Scale. *Int J Rehabil Res*. 2016; 39(1):36–41. <https://doi.org/10.1097/MRR.000000000000141> PMID: [26544627](#)
39. Stratford PW, Kennedy D, Pagura SMC, Gollish JD. The relationship between self-report and performance-related measures: Questioning the content validity of timed tests. *Arthritis Care Res*. 2003; 49(4):535–40. <https://doi.org/10.1002/art.11196> PMID: [12910560](#)
40. Naal F, Impellizzeri F, Torca S, Wellauer V, Leunig M, von Eisenhart-Rothe R. The German Lower Extremity Functional Scale (LEFS) is reliable, valid and responsive in patients undergoing hip or knee replacement. *Qual Life Res*. 2014; 1–6. <https://doi.org/10.1007/s11136-014-0777-6> PMID: [25108549](#)
41. Negahban H, Hessam M, Tabatabaei S, Salehi R, Sohani SM, Mehravar M. Reliability and validity of the Persian lower extremity functional scale (LEFS) in a heterogeneous sample of outpatients with lower limb musculoskeletal disorders. *Disabil Rehabil*. 2014; 36(1):10–5. <https://doi.org/10.3109/09638288.2013.775361> PMID: [23530691](#)
42. Cacchio A, De Blasis E, Necozone S, Rosa F, Riddle DL, di Orio F, et al. The Italian version of the Lower Extremity Functional Scale was reliable, valid, and responsive. *J Clin Epidemiol*. 2010; 63(5):550–7. <https://doi.org/10.1016/j.jclinepi.2009.08.001> PMID: [19913388](#)

43. Lin C-WC, Moseley AM, Refshauge KM, Bundy AC. The Lower Extremity Functional Scale Has Good Clinimetric Properties in People With Ankle Fracture. *Phys Ther*. 2009; 89(6):580–8. <https://doi.org/10.2522/ptj.20080290> PMID: 19423644
44. Cruz-Díaz D, Lomas-Vega R, Osuna-Pérez MC, Hita-Contreras F, Fernández AD, Martínez-Amat A. The Spanish lower extremity functional scale: A reliable, valid and responsive questionnaire to assess musculoskeletal disorders in the lower extremity. *Disabil Rehabil*. 2014; 36(23):2005–11. <https://doi.org/10.3109/09638288.2014.890673> PMID: 24593163
45. Pan S-L, Liang H-W, Hou W-H, Yeh T-S. Responsiveness of SF-36 and Lower Extremity Functional Scale for assessing outcomes in traumatic injuries of lower extremities. *Injury*. 2014; 45(11):1759–63. <https://doi.org/10.1016/j.injury.2014.05.022> PMID: 24938677
46. Abbott JH, Schmitt J. Minimum Important Differences for the Patient-Specific Functional Scale, 4 Region-Specific Outcome Measures, and the Numeric Pain Rating Scale. *J Orthop Sports Phys Ther*. 2014; 44(8):560–4. <https://doi.org/10.2519/jospt.2014.5248> PMID: 24828475
47. Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. *J Clin Epidemiol*. 2006; 59(1):45–52. <https://doi.org/10.1016/j.jclinepi.2005.03.018> PMID: 16360560
48. Williams VJ, Piva SR, Irrgang JJ, Crossley C, Fitzgerald GK. Comparison of Reliability and Responsiveness of Patient-Reported Clinical Outcome Measures in Knee Osteoarthritis Rehabilitation. *J Orthop Sports Phys Ther*. 2012; 42(8):716–23. <https://doi.org/10.2519/jospt.2012.4038> PMID: 22402677
49. René F, Casimiro L, Tremblay M, Brosseau L, Chea P, Létourneau L, et al. Fiabilité test retest et validité de construit de la version française de L'Échelle fonctionnelle des membres inférieurs (ÉFMI), partie II. *Physiother Can*. 2011; 63(2):249–55. <https://doi.org/10.3138/ptc.2010-12F> PMID: 22379266