



Research

Cross-Cultural Adaptation and Psychometric Validation of the Turkish Version of the American Orthopaedic Foot and Ankle Society Lesser Metatarsophalangeal-Interphalangeal Joint Scale

Amerikan Ortopedik Ayak ve Ayak Bileği Derneği Küçük Parmak Metatarsofalangeal-İnterfalangeal Skalasının Türkçe Uyarlaması ve Psikometrik Geçerlik-Güvenirlik Analizi

 Ali Can Koluman¹,  Ebru Aloğlu Çiftçi^{2,3},  Mehmet Utku Çiftçi⁴,  Yasemin Şahbaz⁵,  Vedat Öztürk¹,  Altuğ Duramaz¹,  Emre Baca¹,  Nezih Ziroğlu⁶

¹University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital, Clinic of Orthopedics and Traumatology, İstanbul, Türkiye

²İstanbul Okan University Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, İstanbul, Türkiye

³İstinye University Institute of Graduate Education, Department of Physiotherapy and Rehabilitation, İstanbul, Türkiye

⁴University of Health Sciences Türkiye, Sultan 2. Abdülhamid Han Training and Research Hospital, Clinic of Orthopedics and Traumatology, İstanbul, Türkiye

⁵Beykent University Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, İstanbul, Türkiye

⁶Acıbadem Mehmet Ali Aydınlar University Faculty of Medicine, Department of Orthopedics and Traumatology, İstanbul, Türkiye

ABSTRACT

Objective: Lesser toe disorders can cause significant functional impairment and pain, requiring reliable tools for outcome assessment. The American Orthopaedic Foot and Ankle Society (AOFAS) lesser metatarsophalangeal-interphalangeal (MTP-IP) joint scale is a clinician-based instrument frequently used in foot and ankle evaluations, yet no validated Turkish version exists. The aim of this study was to translate, culturally adapt, and evaluate the psychometric properties of the Turkish version of the AOFAS lesser MTP-IP scale.

Methods: The scale was translated following international cross-cultural adaptation guidelines. A total of 43 patients with various lesser-toe pathologies were assessed using the AOFAS lesser MTP-IP, foot and ankle ability measure (FAAM), visual analogue scale, and short form-12 (SF-12). Test-retest reliability was assessed by calculating intraclass correlation coefficients [ICC (2,1)] using a two-way mixed-effects model with absolute agreement; by assessing internal consistency via Cronbach's alpha; and by evaluating agreement using Bland-Altman analysis. Construct validity was tested by correlating AOFAS scores with FAAM and SF-12 subscales. Floor and ceiling effects were also analyzed.

Results: The Turkish version demonstrated excellent test-retest reliability [ICC (2,1)=0.96] and acceptable internal consistency ($\alpha=0.76$). Bland-Altman plots revealed no systematic bias. Strong correlations were observed with FAAM-activities of daily living ($r=0.93$) and FAAM-sports ($r=0.75$), whereas correlations with SF-12 physical component summary ($r=0.34$) and MCS ($r=0.45$) were weak but significant, which is consistent with the hypothesized convergent and divergent validity. A notable ceiling effect was identified in the AOFAS function and alignment domains, consistent with the high functional status and low pain levels reported by participants.

Conclusion: The Turkish adaptation of the AOFAS lesser MTP-IP scale is a reliable and valid instrument for evaluating pain, function, and alignment in patients with lesser toe disorders. Its strong psychometric performance supports its use in both clinical and research settings, although the observed ceiling effect should be interpreted in the context of patient characteristics.

Keywords: AOFAS, lesser toe, Turkish validation, reliability, validity, psychometrics, foot and ankle

Address for Correspondence: Ali Can Koluman, MD, University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital, Clinic of Orthopedics and Traumatology, İstanbul, Türkiye
E-mail: md.alicankoluman@gmail.com **ORCID ID:** orcid.org/0000-0002-0191-3229

Cite as: Koluman AC, Aloğlu Çiftçi E, Çiftçi MU, Şahbaz Y, Öztürk V, Duramaz A, et al. Cross-cultural adaptation and psychometric validation of the Turkish version of the American Orthopaedic Foot And Ankle Society lesser metatarsophalangeal-interphalangeal joint scale. Med J Bakirkoy. 2025;21(4):458-467

Received: 22.10.2025

Accepted: 11.11.2025

Publication Date: 18.12.2025



ÖZ

Amaç: Küçük parmak patolojileri, belirgin ağrı ve fonksiyonel kısıtlılığa neden olabilir; bu durum, klinik değerlendirme ve tedavi sonuçlarının izlenmesinde güvenilir ölçüm araçlarının kullanılmasını gerektirir. Amerikan Ortopedik Ayak ve Ayak Bileği Derneği (AOFAS) tarafından geliştirilen lesser metatarsofalangeal-interfalangeal (MTP-IP) skalası, ayak ve ayak bileği değerlendirmelerinde sık kullanılan klinisyen temelli bir ölçektir; ancak bu ölçeğin geçerliliği ve güvenilirliği kanıtlanmış Türkçe bir versiyonu bulunmamaktadır. Bu çalışmanın amacı, AOFAS lesser MTP-IP skalasının Türkçe uyarlamasını yapmak ve psikometrik özelliklerini değerlendirmektir.

Gereç ve Yöntem: Ölçeğin çeviri ve kültürel adaptasyonu uluslararası standartlara uygun olarak gerçekleştirildi. Farklı küçük parmak patolojileri bulunan toplam 43 hasta, AOFAS lesser MTP-IP, ayak ve ayak bileği yeteneği ölçeği (FAAM), görsel analog skala ve kısa form-12 (SF-12) kullanılarak değerlendirildi. Test-retest güvenilirliği sınıf içi korelasyon katsayısı [ICC (2,1)] ile, iç tutarlılık Cronbach alfa katsayısı ile, ölçümler arası uyum ise Bland-Altman analiziyle değerlendirildi. Yapı geçerliği, AOFAS puanlarının FAAM ve SF-12 alt ölçekleriyle olan korelasyonları incelenerek test edildi. Ayrıca tavan ve taban etkileri analiz edildi.

Bulgular: Türkçe versiyon mükemmel test-retest güvenilirliği [ICC (2,1)=0,96] ve kabul edilebilir iç tutarlılık ($\alpha=0,76$) gösterdi. Bland-Altman analizinde sistematik yanlılık saptanmadı. AOFAS toplam puanları FAAM-günlük yaşam aktiviteleri ($r=0,93$) ve FAAM-spor ($r=0,75$) ile güçlü, SF-12 fiziksel ($r=0,34$) ve mental ($r=0,45$) bileşen puanları ile zayıf ancak anlamlı korelasyonlar gösterdi. Bu sonuçlar, önceden tanımlanan yakınsak ve iraksak geçerlik hipotezleriyle uyumluydu. Katılımcıların yüksek fonksiyonel durum ve düşük ağrı düzeyleriyle uyumlu olarak, AOFAS fonksiyon ve dizilim alt ölçeklerinde belirgin bir tavan etkisi gözlemlendi.

Sonuç: AOFAS lesser MTP-IP skalasının Türkçe uyarlaması, küçük parmak patolojilerinde ağrı, fonksiyon ve dizilimi değerlendirmede güvenilir ve geçerli bir ölçektir. Ölçeğin güçlü psikometrik özellikleri, klinik ve araştırma ortamlarında kullanılabilirliğini desteklemektedir. Ancak gözlenen tavan etkisi, örneklemin yüksek fonksiyonel düzeyiyle ilişkilendirilerek yorumlanmalıdır.

Anahtar Kelimeler: AOFAS, küçük parmak, Türkçe geçerlik, güvenilirlik, psikometri, ayak ve ayak bileği

INTRODUCTION

Lesser toe disorders, including Freiberg's disease, metatarsalgia, deformities, and sequelae of trauma, are frequently encountered in orthopaedic practice and lead to pain, altered gait mechanics, and functional limitations. Accurate functional assessment is critical not only for guiding clinical decision-making but also for evaluating treatment efficacy in both conservative and surgical contexts (1,2).

Among the tools used to evaluate foot and ankle function, the American Orthopaedic Foot and Ankle Society (AOFAS) clinical rating systems have historically been among the most widely adopted instruments. The AOFAS lesser metatarsophalangeal-interphalangeal (MTP-IP) joint scale, in particular, is designed to assess pain, functional limitations, and alignment in the lesser toes (3). Despite its long-standing utility and simplicity, the scale has been subject to methodological critique because of its categorical structure, ceiling effects, and lack of patient-reported input (4,5). In recognition of these limitations, the AOFAS has officially withdrawn its endorsement of these clinician-based tools, advocating instead for integrating patient-reported outcome measures (PROMs) into both research and practice (6).

Nevertheless, the AOFAS scales remain widely used because of their clinical familiarity and relevance, particularly in reporting postoperative outcomes (7). Recent methodological frameworks—such as COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN)—emphasize the need to evaluate reliability, measurement error, and construct validity

through rigorous statistical procedures, including intraclass correlation coefficients (ICC), standard error of measurement (SEM), and minimal detectable change (MDC) (8,9).

In Türkiye, validated versions of the AOFAS hindfoot and hallux scales are available (10-13); however, there is no psychometrically validated Turkish adaptation of the AOFAS lesser MTP-IP scale. This gap limits the comparability of clinical data across international studies and restricts the use of standardized instruments in Turkish-speaking populations. Despite the increasing emphasis on methodologically sound, culturally adapted tools in musculoskeletal outcomes research, a critical unmet need remains.

Therefore, this study aimed to translate and culturally adapt the AOFAS lesser MTP-IP scale into Turkish and to evaluate its psychometric properties in patients with lesser toe disorders. We hypothesized that the Turkish version would demonstrate strong reliability, internal consistency, and construct validity, supporting its use in both clinical practice and multicenter research efforts.

METHODS**Study Design and Participants**

This cross-sectional methodological study was conducted at a tertiary-level orthopedic outpatient clinic between January and June 2024. Hospital records were used to identify patients who had been treated for a disorder affecting the lesser toe. These patients were invited to attend follow-up visits. Patients who met the eligibility criteria had received treatment within the last two to four

years. A total of 43 patients diagnosed with pathologies of the lesser toes (e.g., Freiberg disease, deformities, or sequelae of trauma) were enrolled (Figure 1). Although the COSMIN guidelines recommend at least 50 participants for an adequate assessment of reliability and validity, previous methodological reports suggest that 5-10 participants per item may be acceptable for focused validation studies (14). Our participant-to-item ratio of 5.4, combined with an ICC power >0.99 , indicates adequate precision despite the modest sample size. For validity studies, it is recommended to include at least five patients per item (14). The AOFAS lesser MTP-IP scale consists of eight items, and our study included 43 patients, corresponding to 5.38 patients per item, thus meeting the recommended sample size criterion. Inclusion criteria included age ≥ 18 years, literacy in Turkish, and clinical or radiographic evidence of a lesser-toe disorder. Patients with systemic neurological conditions, prior major foot surgery, or cognitive impairment were excluded.

This study was approved by the Clinical Research Ethics Committee of University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital (approval no: 2022-06-05, date: 21.03.2022). Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki.

Translation and Cultural Adaptation

The AOFAS lesser MTP-IP scale was translated into Turkish following the five-step process outlined by Beaton et al. (15): (1) forward translation by two bilingual experts; (2) synthesis of translations; (3) back translation by two native English speakers; (4) expert committee review; and (5) pretesting with 20 patients to assess clarity and comprehension. Minor linguistic adjustments were made based on participant feedback to enhance cultural relevance. The translators' professional backgrounds (medical vs. non-medical) and

linguistic expertise were reported to ensure conceptual equivalence and transparency.

Outcome Measurements

AOFAS Lesser MTP-IP Scale (Turkish Version)

The AOFAS lesser MTP-IP scale is a clinician-administered tool designed to evaluate outcomes in patients with disorders of the lesser toes (3). It comprises three domains: pain (40 points), function (45 points), and alignment (15 points), with a total possible score of 100, where higher scores indicate better clinical status. The functional domain includes items related to activity limitations, footwear requirements, and mobility, whereas alignment is rated by the clinician based on joint positioning.

In this study, the original English version of the AOFAS lesser MTP-IP scale was translated and culturally adapted into Turkish using a standardized methodology (Table 1) (15). This Turkish version was administered at two time points, two weeks apart, to assess reliability and construct validity.

Visual Analogue Scale (VAS)

Pain intensity was assessed using a 10-centimeter VAS, which is a validated and widely used unidimensional measure of pain (16). Patients were asked to mark their level of pain on a horizontal line ranging from 0 (no pain) to 10 (worst imaginable pain). VAS scores were recorded in three conditions: at rest, during activity, and at night. This enabled multidimensional pain profiling and facilitated comparison with the AOFAS pain domain. The activity condition refers to pain experienced during daily activities, such as walking, climbing stairs, or performing occupational tasks.

Foot and Ankle Ability Measure

The foot and ankle ability measure (FAAM) is a patient-reported outcome measure (PROMs) developed to assess

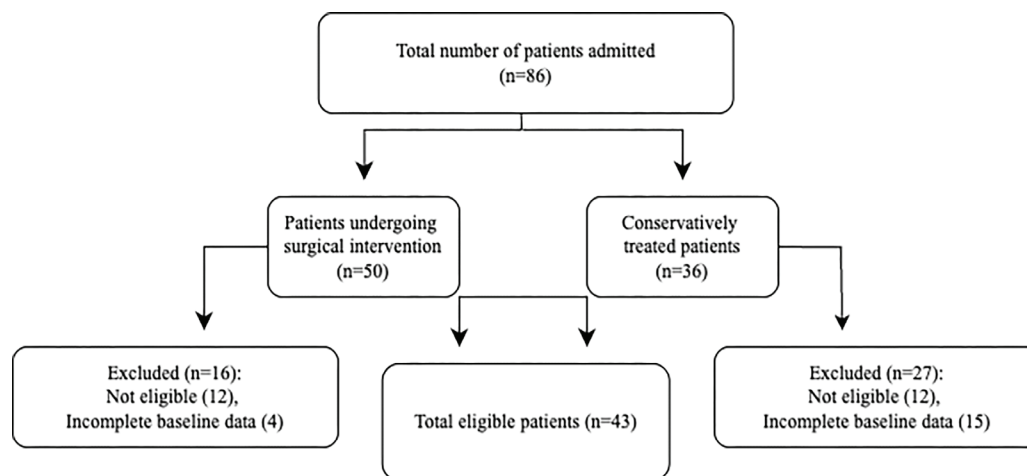


Figure 1. Flowchart. The number of patients in each particular group is shown in brackets

physical function in individuals with musculoskeletal disorders/lesser-extremity (17). It includes two subscales:

- Activities of daily living (ADL): 21 items,
- Sports: 8 items.

Each item is scored on a 5-point Likert scale (0=unable to do; 4=no difficulty), and raw scores are transformed into percentage scores (0-100), with higher scores reflecting greater functional ability. The Turkish version of the FAAM has been validated and shown to have high internal consistency and construct validity (18).

Short Form-12 (SF-12) Health Survey

SF-12 is a generic measure of health-related quality of life derived from the original 36-item SF-36 questionnaire (19). It yields two-component summary scores:

- Physical component summary (PCS),
- Mental component summary (MCS).

These scores are standardized [mean=50, standard deviation (SD)=10], with higher scores indicating better self-perceived health. Although not specific to the foot and ankle, the SF-12 provides insight into general health status and permits evaluation of divergent validity when used alongside region-specific measures such as the AOFAS or FAAM (20).

All instruments were administered in Turkish; validated Turkish versions of the FAAM and SF-12 were used to ensure linguistic and conceptual equivalence. The SF-12 was used to test divergent validity, reflecting general health rather than region-specific function. We hypothesized strong correlations with FAAM subscales and weaker correlations with SF-12 PCS and MCS, consistent with COSMIN recommendations for hypothesis-driven construct validation. The Turkish version validated by Soylu and Kütük (21) was used.

Statistical Analysis

All statistical analyses were performed using IBM SPSS statistics, version 27.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were used to summarize demographic and clinical characteristics. Frequencies and percentages were reported for categorical variables; means and standard deviations were reported for continuous variables. The normality of continuous data was evaluated using the Shapiro-Wilk test.

An ICC value of 0.75 was used as the minimum acceptable threshold for reliability in accordance with established guidelines (22). The observed ICC of 0.96 yielded a post-hoc power of approximately 1.00 ($n=43$, $\alpha=0.05$),

confirming sufficient statistical power to detect excellent test-retest reliability.

Internal consistency of the Turkish version of the AOFAS lesser MTP-IP scale was assessed using Cronbach's alpha. Values between 0.70 and 0.95 were considered acceptable (9,23).

Test-retest reliability was assessed by calculating ICC using a two-way mixed-effects model with absolute agreement [ICC (2,1)]. ICC values were interpreted as follows: <0.40 =poor, $0.40-0.59$ =fair, $0.60-0.74$ =good, and ≥ 0.75 =excellent.

Agreement was evaluated using the SEM and the MDC95. The ICC was applied to compute the SEM, which reflects the precision of the measurement. SEM was determined by multiplying the SD of the scores by the square root of (1-ICC). The MDC95 represents the smallest change that exceeds the threshold of measurement error. To estimate the MDC at the 95% confidence interval (CI) level (MDC 95%), the SEM was multiplied by 1.96 and then by the square root of 2 (22).

Construct validity was evaluated using Spearman's rank correlation coefficients between AOFAS total scores and scores on external instruments (FAAM, VAS, SF-12). Correlation strength was interpreted as small ($0.10-0.29$), moderate ($0.30-0.49$), strong ($0.50-0.69$), and very strong (≥ 0.70), according to Cohen's criteria.

Agreement between test and retest scores was further visualized using Bland-Altman plots, which graphically display the difference between paired measurements against their mean. Systematic bias and 95% limits of agreement (LoA) ($\text{LoA: mean} \pm 1.96 \times \text{SD of differences}$) were examined to determine interchangeability between repeated assessments.

Floor and ceiling effects were evaluated by calculating the proportion of patients scoring the minimum (0-10) or maximum (90-100) possible scores at baseline. Effects were considered present if more than 15% of participants achieved either extreme (24). A p-value of <0.05 was considered statistically significant.

RESULTS

Participant Characteristics

A total of 43 patients (mean age: 38.60 ± 18.70 years; 74.4% female) were included. The most common diagnoses were Freiberg's disease (65.1%), lesser toe fractures (27.9%), and deformities (7.0%). Surgical intervention was performed in 79.1% of cases, and 76.7% of patients reported normal functional status at baseline. Detailed demographic and clinical characteristics are presented in Table 2.

Reliability and Internal Consistency

The Turkish version of the AOFAS lesser MTP-IP scale demonstrated excellent test-retest reliability. The ICC (2,1) for the total score was 0.96 (95% CI: 0.95-0.98), with similarly high values observed for the subscales (pain: 0.96; function:

0.94; alignment: 0.89). Internal consistency, as measured by Cronbach's alpha, was acceptable ($\alpha=0.76$ for the total score and the function subscale), indicating a coherent item structure (Table 3).

Table 1. AOFAS lesser MTP-IP joint scale-Turkish adaptation (AOFAS küçük parmak MTP-IP ekleme ölçeği Türkçe uyarlaması)

Ağrı (40 puan)	Puan
Hiç yok	40
Hafif ve nadiren	30
Orta derecede ve her gün	20
Ciddi ve neredeyse her zaman	0
Fonksiyon (45 puan)	
Aktivite kısıtlılıkları, destek ihtiyacı	
Kısıtlılık yok	10
Günlük yaşam aktivitelerinde kısıtlılık yok (iş sorumlulukları gibi), eğlence aktivitelerinde kısıtlılık	7
Günlük yaşam aktivitelerinde ve eğlence aktivitelerinde kısıtlılık	4
Günlük yaşam aktivitelerinde ve eğlence aktivitelerinde ciddi kısıtlılık	0
Ayakkabı gereksinimleri	
Modaya uygun ve geleneksel ayakkabı, tabanlık gerekmez	10
Rahat ayakkabı, tabanlık gerekir	5
Modifiye edilmiş ayakkabı veya breys	0
MTP ekleme hareketi (dorsifleksiyon ve plantarfleksiyon)	
Normal veya çok hafif kısıtlılık ($\geq 75^\circ\text{C}$)	10
Orta derecede kısıtlılık (30°C - 74°C)	5
Ciddi kısıtlılık ($<30^\circ\text{C}$)	0
IF ekleme hareketi (plantarfleksiyon)	
Kısıtlılık yok	5
Ciddi kısıtlılık ($<10^\circ\text{C}$)	0
MTP-IP stabilitesi (tüm yönlerde)	
Stabil	5
Tamamen instabil veya disloke	0
Nasır oluşumu (küçük parmak MTP-IP ekleme)	
Nasır yok veya semptom yok	5
Nasır var ve semptomatik	0
Dizilim (15 puan)	
İyi, küçük parmak dizilimi iyi	15
Orta, küçük parmakta bir dereceye kadar kötü dizilim, semptom yok	8
Kötü, belirgin derecede kötü dizilim, semptom var	0
Toplam puan	100
AOFAS: Amerikan Ortopedik Ayak ve Ayak Bileği Derneği, MTP-IP: Metatarsophalangeal-interphalangeal	

Table 2. Demographic and clinical characteristics (n=43)

Characteristics	Mean \pm SD
Age, (years)	38.60 \pm 18.70
Hospital stay (days)	1.70 \pm 1.65
Number of comorbidities	0.21 \pm 0.63
Sex, n (%)	
Male	11 (25.6%)
Female	32 (74.4%)
Education, n (%)	
Secondary education	21 (48.8%)
Higher education	22 (51.2%)
Affected side, n (%)	
Right	22 (51.2%)
Left	21 (48.8%)
Surgery on lesser toe, n (%)	
Yes	34 (79.1%)
No	9 (20.9%)
Diagnosis, n (%)	
Freiberg's disease	28 (65.1%)
Lesser-toe fracture	12 (27.9%)
Lesser-toe deformity	3 (7.0%)
Smoking status, n (%)	
Non-smokers	27 (62.8%)
Smokers	16 (37.2%)
Functional status	
Normal	33 (76.7%)
Nearly normal	6 (14.4%)
Abnormal	4 (9.3%)
Visual analogue scale	
Rest	1.00 \pm 1.27
Activity	1.14 \pm 1.40
Night	1.00 \pm 1.34
FAAM	
FAAM-ADL	93.05 \pm 13.48
FAAM-Sports	88.49 \pm 17.64
Short Form-12	
Physical component summary	53.01 \pm 8.72
Mental component summary	55.95 \pm 6.34
FAAM: Foot and Ankle Ability Measure, ADL: Activities of daily living scale, Sports: Sports scale, SD: Standard deviation	

Measurement Error and Agreement

The SEM and MDC95 were calculated as 1.80 and 5.00 points, respectively, for the AOFAS total score. Bland-Altman analysis showed mean differences between test-retest scores (T1, T2) of -0.65 (LoA: -7.65 to 6.35) for AOFAS total; 0.42 (LoA: -1.54 to 2.38) for VAS total; -0.02 (LoA: -7.25 to 5.34) for SF-12 PCS; 0.54 (LoA: -4.82 to 5.90); 0.07 (LoA: -16.05 to 16.19) for FAAM-sports; and -0.36 (LoA: -7.47 to 6.75) for FAAM-ADL. There was no evidence of systematic bias, and the limits of agreement were narrow, supporting the scales' reproducibility (Figure 2). The vast majority of participants fell within ± 1.96 SD of the mean.

Internal Consistency

Cronbach's alpha was 0.76 for both the total AOFAS scale and the function subscale, indicating acceptable internal consistency and item coherence within the domains.

Construct Validity

Strong negative correlations were found between the AOFAS total score and VAS scores for pain at rest ($r=-0.81$), during activity ($r=-0.75$), and at night ($r=-0.82$), all statistically

significant ($p<0.001$). AOFAS scores were also strongly correlated with FAAM-ADL ($r=0.93$) and FAAM-sports ($r=0.75$), demonstrating convergent validity. Moderate correlations were observed with SF-12 PCS ($r=0.34$) and MCS ($r=0.45$), supporting divergent validity (Table 4).

Floor and Ceiling Effects

A ceiling effect was observed in the AOFAS total score, with 60.5% of participants achieving the maximum score. This was most prominent in the alignment and pain domains (90.7% and 76.7%, respectively). Similar ceiling effects were found in the FAAM-ADL (53.5%) and FAAM-sports (41.9%) subscales (Table 3). In contrast, VAS scores exhibited notable floor effects, with approximately half of participants reporting no pain. SF-12 scores did not display significant floor or ceiling effects, though both PCS and MCS subscales were skewed toward higher values, indicating good perceived health.

Importantly, the observed ceiling effect in AOFAS scores aligns with low pain levels and high functional status among participants, suggesting it reflects favorable clinical status rather than a measurement limitation.

Table 3. Test-retest reliability of the Turkish version of the AOFAS lesser MTP-IP scale

		First assessment	Second assessment	Test-retest reliability	SEM	MDC	Floor-ceiling (%)
		Mean \pm SD (95% CI) (n=43)		ICC (95% CI)			
AOFAS lesser MTP-IP-T scale items	Pain	36.97 \pm 5.99 (35.13-38.82)	36.74 \pm 6.06 (34.88-38.61)	0.96 (0.94-0.98)	0.76	2.11	0-76.7
	Function	41.04 \pm 5.59 (39.32-42.76)	40.62 \pm 5.85 (38.82-42.42)	0.93 (0.87-0.96)	1.07	2.97	0-60.5
	Alignment	14.34 \pm 2.05 (13.71-14.98)	14.35 \pm 2.05 (13.72-14.98)	1	0.00	0.00	0-90.7
	Total score	92.21 \pm 12.58 (88.34-96.08)	91.56 \pm 12.91 (87.58-95.53)	0.96 (0.92-0.97)	1.80	5.00	0-60.5
VAS	Rest	1.00 \pm 1.27 (0.61-1.39)	1.58 \pm 1.73 (1.04-2.11)	0.60 (0.37-0.76)	0.74	2.06	48.8-0
	Activity	1.13 \pm 1.40 (0.71-1.57)	1.58 \pm 1.46 (1.12-2.03)	0.65 (0.43-0.79)	0.66	1.83	46.5-0
	Night	1.00 \pm 1.34 (0.59-1.41)	1.26 \pm 1.63 (0.75-1.76)	0.75 (0.59-0.85)	0.55	1.54	51.2-0
FAAM	ADL	93.05 \pm 13.48 (88.90-9.21)	92.69 \pm 12.99 (88.69-96.69)	0.96 (0.93-0.97)	1.82	5.06	0-53.5
	Sports	88.48 \pm 17.64 (83.06-93.92)	88.56 \pm 15.59 (83.76-93.36)	0.87 (0.78-0.93)	4.23	11.73	0-41.9
SF-12	PCS	53.01 \pm 8.72 (50.33-55.70)	52.99 \pm 7.95 (50.54-55.43)	0.90 (0.82-0.94)	1.88	5.22	0-0
	MCS	55.95 \pm 6.34 (54.00-57.90)	56.50 \pm 5.80 (54.71-58.28)	0.89 (0.82-0.94)	1.40	3.89	0-0

AOFAS lesser MTP-IP-T: Turkish version of American Orthopaedic Foot and Ankle Society lesser metatarsophalangeal-interphalangeal joint scale, VAS: Visual analogue scale, FAAM-ADL: Foot and Ankle Ability Measure-activities of daily living scale, SF-12 PCS: Short Form-12 physical component summary, SF-12 MCS: Short Form-12 mental component summary, ICC: Intraclass correlation coefficient, CI: Confidence intervals, SEM: Standard error measurement, MDC: Minimal detectable change, SD: Standard deviation

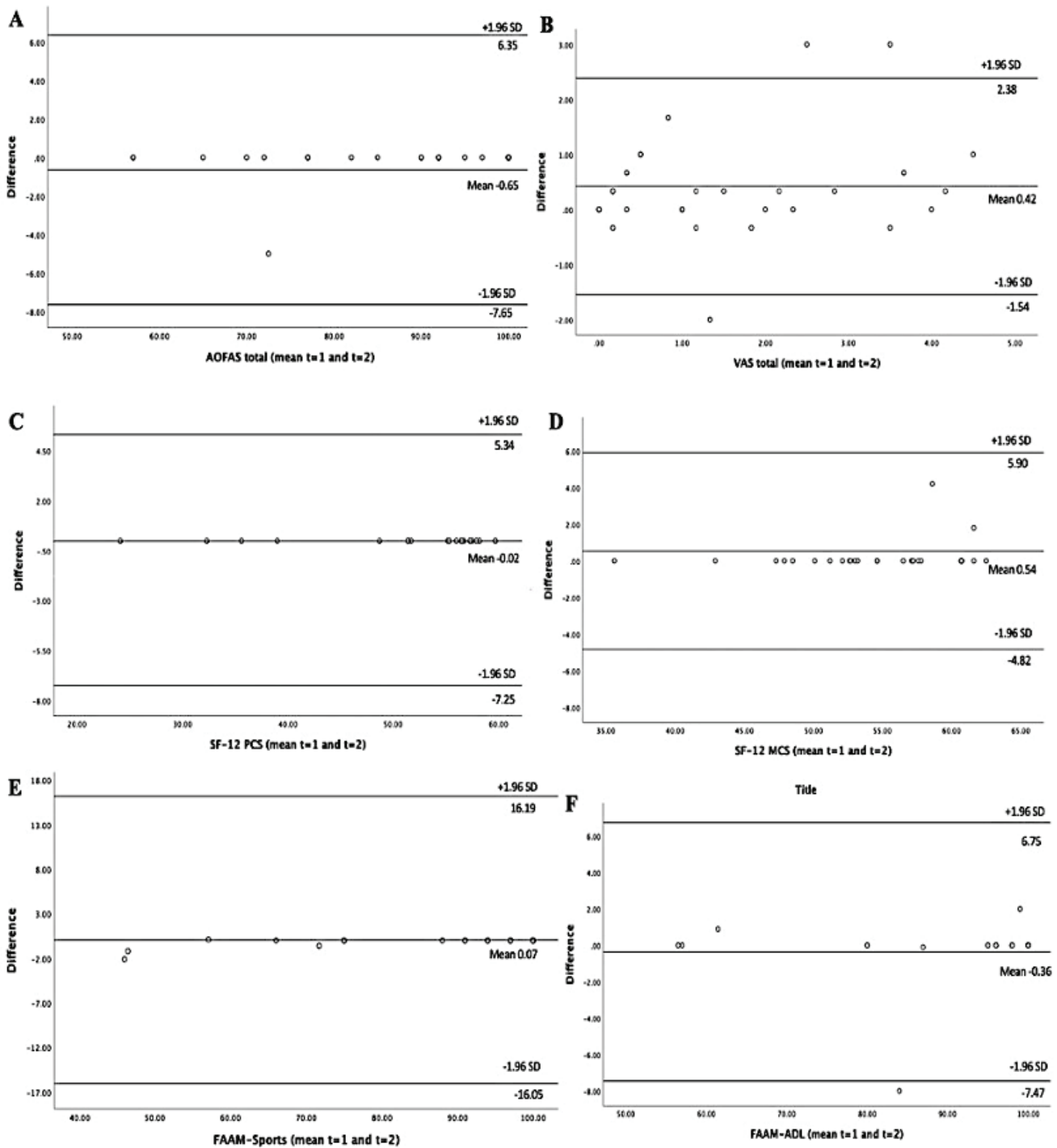


Figure 2. Bland-Altman plots for AOFAS lesser MTP-IP (A), VAS (B), SF-12 PCS (C), SF-12 MCS (D), FAAM-Sports (E) and FAAM-ADL (F) scores in patients; change scores were calculated from time point 1 (t=1) to time point 2 (t=2). VAS (mean of rest, activity and night values). The solid central line represents the mean difference. The upper and lower solid lines represent the 95% limits of agreement, calculated as the mean difference ± 1.96 standard deviation

AOFAS: American Orthopaedic Foot and Ankle Society, MTP-IP: Metatarsophalangeal-interphalangeal, VAS: Visual analogue scale, FAAM: Foot and Ankle Ability Measure, ADL: Activities of daily living scale, Sports: Sports scale, SF-12 PCS: Short Form-12 physical component summary, MCS: Mental component summary

Table 4. Construct and convergent/divergent validity of the Turkish version of the AOFAS lesser MTP-IP scale

	AOFAS lesser MTP-IP-T scale			
	Pain	Function	Alignment	Total
	r (p)			
VAS				
Rest	-0.70 (0.001)**	-0.81 (0.001)**	-0.35 (0.02)*	-0.81 (0.001)**
Activity	-0.66 (0.001)**	-0.73 (0.001)**	-0.33 (0.02)*	-0.75 (0.001)**
Night	-0.68 (0.001)**	-0.82 (0.001)**	-0.32 (0.03)*	-0.82 (0.001)**
FAAM				
ADL	0.76 (0.001)**	0.91 (0.001)**	0.51 (0.001)**	0.93 (0.001)**
Sports	0.65 (0.001)**	0.74 (0.001)**	0.49 (0.001)**	0.75 (0.001)**
Short Form-12				
Physical component summary	0.51 (0.001)**	0.33 (0.02)*	0.26 (0.08)	0.34 (0.02)*
Mental component summary	0.33 (0.02)*	0.42 (0.005)**	0.48 (0.001)**	0.45 (0.002)**
Spearman correlation test: *: p<0.05, **: p<0.01. AOFAS lesser MTP-IP-T scale: Turkish version of American Orthopaedic Foot and Ankle Society lesser metatarsophalangeal-interphalangeal joint scale, VAS: Visual analogue scale, FAAM: Foot and Ankle Ability Measure, ADL: Activities of daily living scale, Sports: Sports scale				

DISCUSSION

This study provides compelling evidence that the Turkish adaptation of the AOFAS lesser MTP-IP scale is a reliable and valid instrument for evaluating pain, function, and alignment in patients with lesser toe pathologies. With excellent test-retest reliability [ICC (2,1)>0.9], acceptable internal consistency, and strong correlations with region-specific (FAAM) and general health (SF-12) instruments, the scale demonstrates robust construct validity and temporal stability key criteria for outcome measures used in clinical research and practice.

Despite the AOFAS scales no longer being officially endorsed because of concerns about responsiveness and lack of patient-centeredness (5,6), their use remains widespread in clinical orthopaedics. This paradox underscores the need for well-validated, culturally adapted versions that enable comparability across studies and acknowledge the inherent limitations of the tool. Although the AOFAS has withdrawn endorsement of its legacy clinician-based scores, these instruments remain in extensive clinical use. Their cultural validation enables comparisons across historical datasets and ongoing multicenter studies, ensuring continuity in outcomes research.

One of the novel contributions of this study is the inclusion of MDC95 and SEM, which provide estimates of measurement precision beyond random error; however, they do not indicate clinically meaningful change (minimal clinically important difference). In addition, presenting Bland-Altman analysis, SEM, MDC95, and floor-ceiling effect results

for all administered instruments further strengthens the comprehensiveness of our evaluation. Applying Bland-Altman analyses further enhances methodological rigor by demonstrating a low systematic bias and good agreement between repeated assessments an aspect often overlooked in prior validation studies. Mean differences close to zero and the majority of data points falling within the ± 1.96 SD limits supported reproducibility. Narrow limits of agreement indicated high reliability for the AOFAS, VAS, SF-12 PCS, SF-12 MCS, and FAAM-ADL scales, whereas wider limits for the FAAM-sports scale suggested greater interindividual variability.

A particularly noteworthy finding was the pronounced ceiling effect observed in the AOFAS pain and alignment domains. The substantial ceiling effect likely reflects the predominantly high-functioning postoperative status of the participants rather than a limitation of the scale itself. Nevertheless, this may restrict the scale's responsiveness in populations with more severe symptoms or functional impairment. Although traditionally viewed as a psychometric limitation, this effect was contextually consistent with the clinical status of the sample: nearly half of the participants reported no pain across VAS domains, and substantial proportions of participants reached maximum FAAM scores. These results parallel those reported by Whittaker et al. (25) and Ponkilainen et al. (26), who noted similar ceiling effects in AOFAS midfoot evaluations effects which could not be fully explained by other outcome metrics. In this light, the ceiling effect in our study appears to reflect favorable clinical outcomes rather than inadequate scale sensitivity.

Nevertheless, some limitations inherent to the AOFAS scoring structure must be acknowledged. Its categorical response format may reduce granularity and responsiveness, particularly in high-functioning or postoperative populations. Future iterations of the scale may benefit from more continuous scaling or the integration of patient input, aligning with broader trends in outcome measurement science.

In this context, the complementary use of PROMs is increasingly essential. Instruments such as FAAM and frameworks like PROMIS offer improved responsiveness, broader biopsychosocial coverage, and in the case of PROMIS, adaptive digital delivery. The integration of PROMIS with clinician-based scales has been shown to enhance the precision and patient-centeredness of musculoskeletal outcome measurement (27,28).

Furthermore, compared with other language versions of the AOFAS scale—including Persian, Italian, and Arabic adaptations—this study distinguishes itself by incorporating SEM, MDC95, and Bland-Altman analyses, which are rarely reported (29-31). These enhancements align with COSMIN standards for outcome measure validation (8,9) and address important methodological gaps in the existing literature.

Study Limitations

Strengths of the study include rigorous adherence to international guidelines for cross-cultural adaptation (15), the use of multiple comparator instruments, and comprehensive statistical analysis. However, the study is limited by its modest sample size, which precluded subgroup analysis by diagnosis or treatment type. Although COSMIN recommendations suggest a minimum sample size of ≥ 50 for reliability studies, our sample included 43 participants. This limitation should be acknowledged, and future studies with larger, more heterogeneous cohorts are warranted to confirm the stability and generalizability of these results. However, the ICC value of 0.96 corresponded to a post-hoc statistical power of >0.99 , indicating that the precision of the reliability estimate was adequate despite the smaller sample size. Additionally, the scale's responsiveness to clinical change over time was not assessed and warrants further longitudinal research. The pronounced ceiling effects, while reflecting high functional recovery, may limit sensitivity in patients with more severe dysfunction and should be considered a psychometric constraint. Additionally, the patient cohort consisted predominantly of postoperative individuals with high functional recovery, which may limit the applicability of the findings to patients with more severe deformity or persistent pain.

CONCLUSION

The Turkish version of the AOFAS lesser MTP-IP scale is a psychometrically robust and clinically practical tool for evaluating lesser toe disorders. The observed ceiling effect, rather than undermining the scale's utility, reflects the functional recovery and low symptom burden of the study population. By providing SEM and MDC95 values and demonstrating strong construct validity, this study enhances the interpretability and applicability of the AOFAS scale in both clinical and research settings. Future studies should investigate the scale's responsiveness and explore its integration with PROMIS-based instruments to support a more holistic and precise approach to musculoskeletal outcome assessment. Future studies in larger, more heterogeneous populations are warranted to confirm the measurement properties and responsiveness of the Turkish AOFAS lesser MTP-IP scale.

ETHICS

Ethics Committee Approval: This study was approved by the Clinical Research Ethics Committee of University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital (approval no: 2022-06-05, date: 21.03.2022).

Informed Consent: Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki.

FOOTNOTES

Authorship Contributions

Surgical and Medical Practices: V.Ö., A.D., E.B., Concept: A.C.K., Y.Ş., N.Z., Design: A.C.K., Y.Ş., N.Z., Data Collection or Processing: A.C.K., M.U.Ç., V.Ö., Analysis or Interpretation: E.A.Ç., Y.Ş., A.D., E.B., Literature Search: A.C.K., E.A.Ç., Writing: A.C.K., M.U.Ç., N.Z.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declare that this study received no financial support.

REFERENCES

1. Coughlin MJ, Mann RA, Saltzman CL. Surgery of the foot and ankle. Elsevier Mosby; 2006.
2. Carmont MR, Rees RJ, Blundell CM. Current concepts review: Freiberg's disease. Foot Ankle Int. 2009;30:167-76.
3. Kitaoka HB, Alexander IJ, Adelaar RS, Nunley JA, Myerson MS, Sanders M. Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. Foot Ankle Int. 1994;15:349-53.

4. Guyton GP. Theoretical limitations of the AOFAS scoring systems: an analysis using Monte Carlo modeling. *Foot Ankle Int.* 2001;22:779-87.
5. Pinsker E, Daniels TR. AOFAS position statement regarding the future of the AOFAS clinical rating systems. *Foot Ankle Int.* 2011;32:841-2.
6. Kitaoka HB, Meeker JE, Phisitkul P, Adams SB Jr, Kaplan JR, Wagner E. AOFAS position statement regarding patient-reported outcome measures. *Foot Ankle Int.* 2018;39:1389-93.
7. Hijji FY, Schneider AD, Pyper M, Laughlin RT. The popularity of outcome measures used in the foot and ankle literature. *Foot Ankle Spec.* 2020;13:58-68.
8. 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;19:539-49.
9. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34-42.
10. Analay Akbaba Y, Celik D, Ogut RT. Translation, cross-cultural adaptation, reliability, and validity of Turkish version of the American Orthopaedic Foot and Ankle Society ankle-hindfoot scale. *J Foot Ankle Surg.* 2016;55:1139-42.
11. Kalaycıoğlu O, Ayanoğlu T, Yılmaz ÖF, Turhan MT. A cross-cultural adaptation and validation of the Turkish version of American Orthopaedic Foot and Ankle Society metatarsophalangeal-interphalangeal scale (AOFAS-MTP-IP) for the hallux. *Jt Dis Relat Surg.* 2023;34:305-14.
12. Ziroglu N, Birinci T, Koluman A, Şahbaz Y, Çiftçi MU, Baca E, et al. Reliability and validity of the Turkish version of the American Orthopaedic Foot and Ankle Society hallux metatarsophalangeal-interphalangeal joint scale. *Foot Ankle Spec.* 2025;18:579-88.
13. Ziroğlu N, Şahbaz Y, Koluman A, Birinci T, Çiftçi MU, Baca E. Cross-cultural adaptation, reliability, and validity of the Turkish version of the American Orthopedic Foot and Ankle Society (AOFAS) midfoot scale. *Med J Bakirkoy.* 2024;20:264-70.
14. Costello AB, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Pract Assess Res Eval.* 2005;10:7.
15. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976).* 2000;25:3186-91.
16. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: visual analog scale for pain (VAS Pain), numeric rating scale for pain (NRS Pain), mcgill pain questionnaire (MPQ), short-form mcgill pain questionnaire (SF-MPQ), chronic pain grade scale (CPGS), short form-36 bodily pain scale (SF-36 BPS), and measure of intermittent and constant osteoarthritis pain (ICOAP). *Arthritis Care Res (Hoboken).* 2011;63(Suppl 11):S240-52.
17. Martin RL, Irrgang JJ, Burdett RG, Conti SF, Van Swearingen JM. Evidence of validity for the foot and ankle ability measure (FAAM). *Foot Ankle Int.* 2005;26:968-83.
18. Çelik D, Malkoç M, Martin R. Evidence for reliability, validity and responsiveness of Turkish foot and ankle ability measure (FAAM). *Rheumatol Int.* 2016;36:1469-76.
19. Ware J Jr, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34:220-33.
20. Ibrahim T, Beiri A, Azzabi M, Best AJ, Taylor GJ, Menon DK. Reliability and validity of the subjective component of the American Orthopaedic Foot and Ankle Society clinical rating scales. *J Foot Ankle Surg.* 2007;46:65-74.
21. Soylu C, Kütük B. Reliability and validity of the Turkish version of SF-12 health survey. *Turk Psikiyatri Derg.* 2022;33:108.
22. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15:155-63. Erratum in: *J Chiropr Med.* 2017;16:346.
23. Nunnally JC, Bernstein IH. *Psychometric theory ed.*, New York; 1994.
24. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res.* 1995;4:293-307.
25. Whittaker GA, Munteanu SE, Roddy E, Menz HB. Measures of foot pain, foot function, and general foot health. *Arthritis Care Res (Hoboken).* 2020;72(Suppl 10):294-320.
26. Ponkilainen VT, Uimonen M, Repo JP, Mattila VM, Haapasalo HH. Validity and internal consistency of the American Orthopaedic Foot and Ankle Society midfoot scale in patients with Lisfranc injury. *Foot Ankle Surg.* 2020;26:523-9.
27. Bernstein DN, Kelly M, Houck JR, Ketz JP, Flemister AS, DiGiovanni BF, et al. PROMIS pain interference is superior vs numeric pain rating scale for pain assessment in foot and ankle patients. *Foot Ankle Int.* 2019;40:139-44.
28. Hung M, Baumhauer JF, Licari FW, Bounsanga J, Voss MW, Saltzman CL. Responsiveness of the PROMIS and FAAM instruments in foot and ankle orthopedic population. *Foot Ankle Int.* 2019;40:56-64.
29. Alhadhoud M, Alsiri N, Alsaffar M, Glazebrook M. Cross-cultural adaptation and validation of an Arabic version of the American Orthopedics Foot and Ankle score (AOFAS). *Foot Ankle Surg.* 2020;26:876-82.
30. Vosoughi AR, Yoosefinejad AK, Dehbarez YS, Kargarshouraki Z, Mahdaviazad H. Evaluating the validity and reliability of the Persian version of American Orthopedic Foot and Ankle Society midfoot scale. *Foot Ankle Spec.* 2024;17:442-50.
31. Leighab M, Vaiuso D, Rava E, Pogliacomini F, Samaila EM, Grassi FA, et al. Translation, cross-cultural adaptation, reliability, and validation of the Italian version of the American Orthopaedic Foot and Ankle Society-metatarsophalangeal-interphalangeal scale (AOFAS-MTP-IP) for the hallux. *Acta Biomed.* 2019;90(12-S):118-26.