Joint Diseases and Related Surgery

Jt Dis Relat Surg
2023;34(2):305-314

ORIGINAL ARTICLE

# 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

Oya Kalaycıoğlu, PhD[1], Tacettin Ayanoğlu, MD[2], Ömer Faruk Yılmaz, MD[2], Mahmut Timur Turhan, MD[2]

[1]Department of Biostatistics and Medical Informatics, Bolu Abant İzzet Baysal University, Bolu, Türkiye
[2]Department of Orthopedics and Traumatology, Bolu Abant İzzet Baysal University, Bolu, Türkiye

Pathophysiologically, hallux valgus (HV) is a common deformity caused by the outward deviation of the big toe and inward deviation of the first metatarsal. It is usually accompanied by a bony prominence described as a "bunion" in the medial part of the first metatarsophalangeal (MTP) joint. It causes various deformities (claw and hammer toe) in the other fingers, as the thumb slides inward due to HV. If HV is left untreated, it shows a progressive nature that affects the daily living activities of patients, leading to foot deformity, and severe pain in the future.[1] Therefore, it is crucial to regularly monitor the progress of HV to improve patients' quality of life and functional ability.

Received: April 05, 2023
Accepted: May 12, 2023
Published online: May 25, 2023

**Correspondence**: Oya Kalaycıoğlu, PhD. Bolu Abant İzzet Baysal Üniversitesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, 14030 Gölköy, Bolu, Türkiye.

E-mail: oyakalaycioglu@ibu.edu.tr

Doi: 10.52312/jdrs.2023.1141

**Citation:** 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(2):305-314. doi: 10.52312/jdrs.2023.1141.



## ABSTRACT

**Objectives:** In this study, we aimed to translate and culturally adapt the American Orthopaedic Foot and Ankle Society (AOFAS) Hallux Metatarsophalangeal (MTP)-Interphalangeal (IP) scale, which is used for the clinical assessment of patients with hallux valgus (HV), into Turkish and to evaluate its validity and reliability.

**Patients and methods:** Between February 2022 and October 2022, a total of 67 patients (18 males, 49 females; mean age: 51.5±15.9 years; range, 18 to 68 years) with HV deformity and able to communicate in Turkish were included. Following the translation of the AOFAS hallux MTP-IP scale into Turkish, its cultural appropriateness was confirmed. Intra-rater and inter-rater reliabilities were assessed by intraclass correlation coefficients (ICCs), using data collected by two orthopedists. Agreement among test-retest evaluations was conducted using the Bland-Altman analysis. The construct validity of the scale was determined by the Manchester-Oxford Foot Questionnaire (MOXFQ) and Short Form Health Survey (SF-36). Content validity was confirmed by the floor/ceiling effects.

**Results:** The Turkish AOFAS hallux MTP-IP had an excellent intra-rater reliability of 0.971. The intra-rater reliability of the pain, function, and alignment subscales ranged from 0.904 to 0.978. The inter-rater reliability was 0.913 for the total score, while ranging from 0.838 to 0.918 for the subscales. The total score of the AOFAS hallux MTP-IP had a high correlation with the physical domains of the MOXFQ and SF-36, while weaker correlations with mental domains were observed. No floor/ceiling effect was observed for the overall Turkish AOFAS hallux MTP-IP.

**Conclusion:** The Turkish translated and culturally adapted AOFAS hallux MTP-IP scale is a valid and reliable measure, ensuring its use in assessing the clinical status of Turkish patients with HV deformity.

***Keywords:*** American Orthopaedic Foot and Ankle Society, hallux, metatarsophalangeal joint.reliability, Turkish, validity.

The prevalence of HV deformity varies in epidemiological studies and has been reported to be between 21 and 70%.[2] Given the high prevalence, a wide range of clinical scoring systems are frequently used for the clinical assessment of patients with HV deformity. These scoring systems provide clinicians with reliable and valid measures of the issues that patients are most concerned about such as pain, function, mobility, and health-related quality of life.[3] The most common rating systems targeting foot problems include, but are not limited to, the Manchester-Oxford Foot Questionnaire (MOXFQ),[4] Foot Function Index (FFI),[5] Foot and Ankle Ability Measure (FAAM),[6] Foot and Ankle Disability Index,[7] and American Orthopaedic Foot and Ankle Society (AOFAS) clinical scoring scales.[8] Developed in 1994, the AOFAS scale is the most popular outcome measure for the evaluation of a variety of foot and ankle procedures and disorders.[9] In particular, the AOFAS for the hallux MTP and interphalangeal (IP) joint scale assists clinicians in the assessment of the clinical progress and quality of life of patients with HV deformity based on both subjective and objective components.

In scientific studies, while using questionnaires originally developed in other countries and languages, it is necessary to perform cross-cultural adaptation and validation in addition to translation.[10] Despite the high prevalence of forefoot deformities in the Turkish population[11] and the wide range of use of AOFAS scales across different countries, the Turkish adaptation of the AOFAS scale for assessing pathologies in hallux MTF-IP joints has not yet been carried out. Therefore, in the present study, we aimed to translate and culturally adapt the AOFAS hallux MTF-IP scale into Turkish and to verify its reliability and validity.

## PATIENTS AND METHODS

### Participants

This single-center, observational study was conducted at Bolu Abant İzzet Baysal University, Department of Orthopedics and Traumatology between February 2022 and October 2022. A total of 67 patients (18 males, 49 females; mean age: 51.5±15.9 years; range, 18 to 68 years) with HV deformity and able to communicate in Turkish were included. For the patient distribution in the study to be homogeneous, only patients with HV were enrolled. The diagnosis was made by orthopedic surgeons based on specific symptoms, physical examination findings, and foot radiographs. Patients with cognitive impairment, mental or psychological problems, systemic diseases such as rheumatoid arthritis, and early postoperative and acute trauma with fractures were excluded from the study.

The sample size was found to be adequate based on the recommendations for the required number of subjects in validity and reliability studies. For the validation study, a minimum of five respondents per item is required.[12] The AOFAS hallux MTP-IP scale consists of eight items; thus, a minimum of 40 patients were considered adequate for the validation study. The sample size estimation for reliability was estimated as 46, based on an acceptable reliability of at least 0.80 with an intraclass correlation coefficient (ICC) specific value of 0.90 at $\alpha$=0.05, with a power of 80% for repeating the measurement twice.[13] These values were determined based on previous studies that assessed the reliability of the AOFAS hallux MTP-IP.[14-17]

### Testing protocol

Prior to the study, approval for the use and translation of the AOFAS hallux MTP-IP scale was obtained from the first author of the original article.[8]

Demographic baseline data of the patients, including age, sex, weight, height, education level, previous surgeries on the forefoot, and the use of any walking aid (e.g., stick, crutch, or walker), were recorded upon arrival. The Turkish AOFAS hallux MTP-IP scale and the previously validated Turkish versions of the MOXFQ[18] and SF-36[19] were implemented by two orthopedists who were trained to read the questions and record the participants' answers via Google forms. The same orthopedist administered the AOFAS hallux MTP-IP scale to the same patients one to three days after the first assessment. We kept the re-assessment period short not to postpone the treatment for the patients.

### Questionnaires

*AOFAS hallux MTP-IP scale:* This scale was developed by Kitoaka et al.[8] in 1994 to determine the degree of HV deformities in individuals with toe disorders. It is an eight-item questionnaire consisting of three subscales: pain (40 points), function (45 points), and alignment (15 points), with a total score ranging from 0 to 100. A score of 100 points indicates no symptoms or impairments, and is possible for a patient with no pain, full function, and good alignment. The scale combines subjective and objective questions to obtain scores. Subjective scores, which are rated by the patient,

include questions regarding pain, activity limitation, and footwear requirements. The objective scores, which are rated by clinicians, include evaluations of the range of motion of the MTP and IP joints, MTP or IP instability, and alignment.

*MOXFQ:* MOXFQ, a patient-reported outcome measure following foot or ankle surgery, comprises three subscales: pain, walking/standing, and social interaction.[4] It consists of 16 questions answered on a five-point Likert scale (each item is scored from 0 to 4, with 4 indicating a very severe condition). Raw subscale scores are converted to a numeric scale from 0 to 100, where 100 denotes the most severe medical condition.

*Short Form Health Survey (SF-36):* The SF-36 questionnaire developed by Ware and Sherbourne[20] in 1992 and is a generic scoring tool used to evaluate the health status of individuals. It consists of eight subscale scores that assesses eight health concepts: physical functioning (PF), role limitations due to physical health (PR), bodily pain (BP), general health perceptions (GH), social functioning (SF), role limitations due to emotional problems, vitality (energy and fatigue) (VT), and mental health (MH). These subscale scores are, then, aggregated to obtain the physical component summary (PCS) score and mental composite component summary (MCS) score. A patient can score a minimum of 0 and a maximum of 100 for each subscale, with higher scores indicating better medical conditions.

**Translation and cultural adaptation**

The Beaton's five-stage cross-cultural adaptation principle was used for the translation and cultural adaptation of the AOFAS hallux MTP-IF scale.[21] In the first stage, two bilingual translators (an orthopedic surgeon and an engineer) whose native language is Turkish produced two independent translations. A synthesis of these translations was then made, and both translators agreed on a translated version. Working with the synthesized Turkish version, the back-translations were produced by two native English speakers, who are fluent in Turkish. These translators were neither aware of the study, nor had a medical background. The review team, consisting of translators and experts in the field, evaluated the two back-translations and achieved equivalence between the original and Turkish versions of the AOFAS hallux MTP-IF. Finally, the translated version was pre-tested in a group of 10 patients from the target setting for cultural adaptation. Feedback was obtained from all patients on whether they understood the questions and were able to provide answers. As "recreational activities" is not a commonly used phrase in Turkish, patients required additional explanation for this phrase. Therefore, "leisure activities" was used instead of "recreational activities" in Item 2. No further changes to the questionnaire were necessary, as none of the patients reported any ambiguities or misunderstandings. Therefore, the Turkish version of the AOFAS questionnaire was considered final (Appendix 1).

**Statistical analysis**

Statistical analysis was performed using the R version 4.0.3 software (R Development Core Team, 2010; www.R-project.org) and run in RStudio (www.rstudio.com).[22] The normality assumption for continuous variables was assessed using the Shapiro-Wilk test. Continuous data were expressed in mean ± standard deviation (SD) or median and interquartile range (IQR), while categorical data were expressed in number and frequency. The reliability and validity measures including internal consistency, inter-rater and intra-rater reliabilities, test-retest agreement, construct validity, and floor/ceiling effects were examined for the AOFAS hallux MTP-IP scale, as described below. A $p$ value of $<0.05$ was considered statistically significant.

**Reliability**

Reliability represents the ability of a scale to provide consistent results when repeated under identical conditions. In this study, the following five elements of reliability were determined: internal consistency, inter-rater reliability, intra-rater reliability, test-retest reliability, and agreement. Internal consistency was assessed using Cronbach's alpha (α), and an α of >0.70 was considered acceptable.[23] The inter-rater and intra-rater reliabilities of the AOFAS score were calculated based on evaluations conducted on the same day with 1-h interval. Inter-rater reliability was calculated using repeated measurements by different examiners, while intra-rater reliability was determined using repeated measurements taken by the same examiner. For test-retest reliability, 36 patients were evaluated again one to three days after the first assessment by the same examiner. To maintain the clinical condition of the patients, no treatment for HV was provided between the test periods. The ICCs were calculated using a two-way random-effects model with a corresponding 95% confidence interval (CI). ICC values greater than 0.70 were deemed as acceptable and were interpreted as excellent (0.81-1.0), very good (0.61-0.80), good (0.41-0.60), fair (0.21-0.40), and poor (0.00-0.20).[24]

The degree of agreement between repeated test-retest measurements of the AOFAS hallux MTP-IP was determined with the standard error of the mean (SEM= SD × (√1-ICC)) and minimal detectable change (MDC= 1.96 × √2 × SEM).[25] In the MDC formula, √2 is used to account for the underlying uncertainty of two measurement time points and 1.96 is the corresponding Z-score value for the 95% CI. To present the relative amount of random measurement error, MDC% was calculated as ($MDC/Mean_{all\ scores}$) × 100%, representing the relative amount of random measurement error. An MDC% <30% was considered acceptable and <10% was deemed excellent; however, these cut-off points were more or less arbitrary.[26] The degree of absolute agreement was also assessed using Bland-Altman plots, which demonstrate within-subject variation. The limits of agreement were defined as the Meanchange ± 1.96 × $SD_{change}$. Zero falling outside these limits indicates systematic differences (bias) in the repeated measurements.

**Validity**

Validity, which refers to the degree to which a measuring tool captures what it claims to measure, is typically assessed using construct and content validity.[27] In this study, convergent and discriminant validity were used to distinguish between the two aspects of construct validity. Convergent validity evaluates how closely the measurement tool is related to other variables that measure a similar construct, whereas discriminant validity examines the relationship between the measurement tool and other unrelated variables.[28] Evidence for convergence validity of the AOFAS hallux MTP-IP was determined by its correlation with the physical component scores of the MOXFQ (walking/standing and pain) and SF-36 (PF, RP, BP, GH). The social interaction component of the MOXFQ and the emotional and mental components of the SF-36 (VT, SF, RE, MH) were used to assess discriminant validity. It is hypothesized that the physical components of the MOXFQ and SF-36 are strongly correlated with the AOFAS-MTP-IP score, whereas a low correlation is expected with the mental components. Due to the non-normal distribution of the data, the Spearman rank (r) correlations were used to measure construct validity. Construct validity was considered acceptable, if the correlation of the scale with an instrument measuring the same construct was >0.50, while correlations with related constructs were higher than those with unrelated constructs.[29]

Content validity was evaluated by the floor/ceiling effect, defined as the proportion of patients with maximum (ceiling) and minimum (floor) scores on the scale to the total number of patients. Floor/ceiling effects were considered present, if 15% of the patients had a score at the lowest or highest limits of the scale.[30]

**TABLE I**
Baseline characteristics of patients (n=67)

| Variables | n | % | Mean±SD | Range |
|---|---|---|---|---|
| Age (year) | | | 51.5±15.9 | 18-68 |
| Sex | | | | |
| Male | 18 | 26.9 | | |
| Female | 49 | 73.1 | | |
| Body mass index (kg/m²) | | | 28.7±5.6 | 18.9-42.5 |
| Education | | | | |
| Primary school | 38 | 56.7 | | |
| Secondary school | 8 | 12.0 | | |
| High school | 10 | 14.9 | | |
| Bachelor or higher | 11 | 16.4 | | |
| Involved foot | | | | |
| Right | 19 | 28.4 | | |
| Left | 24 | 35.8 | | |
| Bilateral | 24 | 35.8 | | |
| Surgery on hallux (yes) | 3 | 4.5 | | |
| Use of walking aid (yes) | 13 | 19.4 | | |

SD: Standard deviation.

**TABLE II**
Summary statistics of the patient-reported outcome measures

| Scales | Mean±SD | Median | Range |
|---|---|---|---|
| AOFAS assessment 1 | 70.6±20.6 | 72 | 10-100 |
| AOFAS assessment 2 | 72.6±15.6 | 72 | 19-100 |
| MOXFQ | | | |
| Walking/standing | 46.0±15.1 | 51.4 | 11.4-74.3 |
| Pain | 41.6±13.9 | 42.0 | 8.0-68.0 |
| Social | 32.1±15.7 | 30.0 | 5.0-65.0 |
| SF-36 | | | |
| Physical component summary | 57.2±18.7 | 55.0 | 20.3-89.7 |
| Mental component summary | 57.9±14.8 | 59.8 | 22.4-87.9 |

SD: Standard deviation; AOFAS: American Orthopaedic Foot and Ankle Society; MOXFQ: Manchester-Oxford Foot Questionnaire; SF-36: 36-Item Short form health survey.

## RESULTS

A total of 67 patients with HV deformity were included in the study. The baseline demographic and clinical characteristics of the patients are shown in Table I. Thirty-six patients (7 males and 29 females, mean age: 49.9±15.6 years; range, 21 to 64 years) participated to the second assessment for examining agreement across repeated measurements. Table II shows the descriptive statistics for the AOFAS hallux MTP-IP scale at baseline and the second administration, as well as other outcome measures.

### Internal consistency

The Cronbach's $\alpha$ coefficient was 0.704 for the AOFAS Hallux MTP-IP total score and 0.706 for the functional domain. Internal consistencies were not assessed for the AOFAS pain and alignment subscales, as they had only one item.

### Intra-rater and Intra-rater reliability

The AOFAS hallux MTP-IP scale has an excellent intra-rater reliability of 0.971 (Table III). The ICCs of the pain, function, and alignment subscores ranged from 0.904 to 0.978, showing good intra-rater reliability for the subscales. The inter-rater reliability was 0.913 for the total score and ranged between 0.838 and 0.918 for the subscales.

### Agreement and test-retest reliability

The level of agreement between repeated measurements was demonstrated using SEM and the

**TABLE III**
Reliability and agreement of Turkish AOFAS hallux MTP-IP

| | Intra-rater reliability | | Inter-rater reliability | | Agreement | | |
|---|---|---|---|---|---|---|---|
| AOFAS hallux MTP-IP (sub)scales | ICC | 95% CI | ICC | 95% CI | SEM | MDC | MDC% |
| Pain | 0.978 | 0.959-0.988 | 0.838 | 0.704-0.914 | 3.5 | 9.7 | 32.0 |
| Function | | | | | | | |
| Activity limitation | 0.973 | 0.950-0.986 | 0.825 | 0.686-0.907 | 1.2 | 3.3 | 38.4 |
| Footwear requirements | 1.000 | | 0.877 | 0.770-0.936 | 1.0 | 2.8 | 38.3 |
| MTP joint motion | 0.816 | 0.679-0.898 | 0.845 | 0.717-0.918 | 1.0 | 2.8 | 31.2 |
| IP joint motion | 0.714 | 0.521-0.838 | 0.724 | 0.521-0.850 | 0.8 | 2.3 | 43.7 |
| MTP-IP stability | 0.782 | 0.624-0.878 | 0.660 | 0.430-0.810 | 0.6 | 1.6 | 30.5 |
| Callus | 1.000 | | 0.930 | 0.869-0.964 | 0.6 | 1.6 | 39.8 |
| Function total | 0.904 | 0.826-0.948 | 0.918 | 0.846-0.957 | 2.5 | 6.8 | 19.8 |
| Alignment | 0.955 | 0.917-0.976 | 0.875 | 0.770-0.934 | 1.5 | 4.1 | 33.9 |
| Total | 0.971 | 0.946-0.985 | 0.913 | 0.836-0.955 | 5.1 | 14.2 | 19.8 |

AOFAS: American Orthopaedic Foot and Ankle Society; MTP: Metatarsophalangeal; IP: Interphalangeal; CI: Confidence interval; ICC: Intraclass correlation coefficient; SEM: Standard error of the mean; MDC: Minimal detectable change.

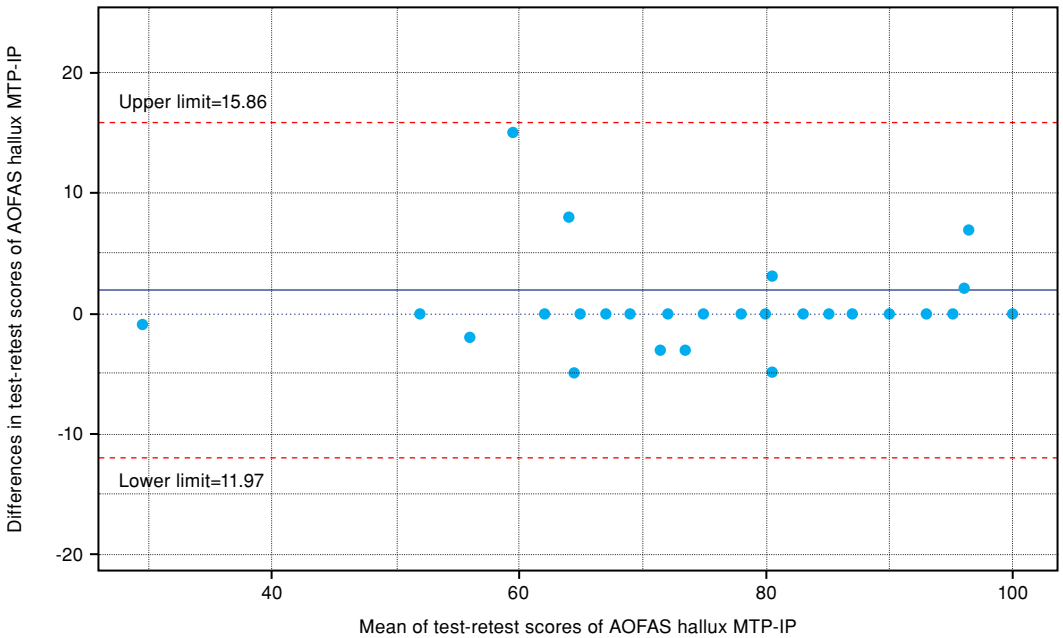


**FIGURE 1.** Brand-Altman plot for assessing the agreement between test-retest measurements of the Turkish AOFAS hallux MTP-IP (Dashed lines indicate the 95% limits of agreement $Mean_{change}$ ±1.96 × $SD_{change}$).
AOFAS: American Orthopaedic Foot and Ankle Society; MTP: Metatarsophalangeal; IP: Interphalangeal.

corresponding MDC (MDC%), as listed in Table III. The MDC was 14.2 (19.8%) for the overall AOFAS hallux MTP-IP score, 9.7 (32.0%) for the pain, 6.8 (19.8%) for the function, and 4.1 (33.9%) for alignment subscores, representing moderate-to-good random measurement error. The Bland-Altman plot (Figure 1) indicated no evidence of systematic bias in test-retest measurements, as zero lies within the 95% limits of agreement for the mean change in AOFAS hallux MTP-IP scores.

**TABLE IV**
Construct validity of Turkish AOFAS Hallux MTP-IP Scale

| | AOFAS hallux MTP-IP sub-scores | | | |
|---|---|---|---|---|
| Sub-scales | Pain | Function | Alignment | Total |
| MOXFQ | | | | |
| Walking/standing | -0.515‡ | -0.547‡ | -0.351‡ | -0.591‡ |
| Pain | -0.654‡ | -0.642‡ | -0.410‡ | -0.732‡ |
| Social | -0.110 | -0.166 | 0.035 | -0.131 |
| SF-36 | | | | |
| Physical functioning | 0.181 | 0.554‡ | 0.342‡ | 0.404‡ |
| Role limitations due to physical health | 0.147 | 0.470‡ | 0.265* | 0.350‡ |
| Bodily pain | 0.542‡ | 0.599‡ | 0.342‡ | 0.620‡ |
| General health | 0.093 | 0.068 | -0.107 | 0.025 |
| Social functioning | 0.317‡ | 0.339‡ | 0.133 | 0.350‡ |
| Role limitations due to emotional problems | 0.011 | 0.433‡ | 0.280* | 0.280* |
| Vitality | 0.051 | 0.082 | -0.065 | 0.059 |
| Mental health | -0.008 | 0.188 | 0.063 | 0.106 |
| Physical component summary | 0.292* | 0.601‡ | 0.334‡ | 0.485‡ |
| Mental component summary | 0.049 | 0.392‡ | 0.159 | 0.254* |

AOFAS: American Orthopaedic Foot and Ankle Society; MTP: Metatarsophalangeal; IP: Interphalangeal; MOXFQ: Manchester-Oxford Foot Questionnaire; SF-36: 36-Item Short form health survey; ‡ Statistical significance at p<0.05 level; * Statistical significance at p<0.01 level. Note that the AOFAS and SF-36 scales are in the opposite direction to the MOXFQ in terms of scores.

**Construct validity**

The total score of the AOFAS hallux MTP-IP was highly correlated with MOXFQ walking/standing (r=-0.591, p<0.001), MOXFQ pain (r=-0.732, p<0.001), and physical component summary of SF-36 (r=0.485, p<0.001). Additionally, the pain and function scores of the AOFAS hallux MTP-IP showed high correlations with the pain and function domains of the MOXFQ and SF-36, respectively (Table IV). These findings suggest a good convergent validity. Weaker correlations were observed with the social domain of the MOXFQ (r=-0.131, p=0.290) and the MH domain of the SF-36 (r=0.106, p=0.393), suggesting good discriminant validity.

**Content validity**

For the first admission, no floor/ceiling effect was present for the overall Turkish AOFAS hallux MTP-IP, as none of the patients who completed the scale had the lowest possible score and only four patients (6%) had the highest possible score. Similarly, the score distribution showed no floor/ceiling effect at the second admission, as none of the patients had the lowest possible score, and only two patients (6%) had the highest possible score.

## DISCUSSION

The AOFAS hallux MTP-IP scale is a scoring system used to evaluate the severity of hallux deformities and related functional impairments.[9] It covers a range of aspects related to hallux-related conditions including pain, function, alignment, and range of motion. This comprehensive assessment can help clinicians to identify specific areas of concern and to tailor treatment plans to address them. Therefore, this scale is widely used by foot and ankle specialists to evaluate the effectiveness of treatment interventions, including surgical procedures and non-surgical therapies. Using a standardized scoring system such as the AOFAS hallux MTP-IP scale, healthcare providers can ensure that patients receive appropriate and effective treatment and that their progress is followed in a consistent and objective manner.

There are several scoring systems in the literature similar to the AOFAS hallux MTP-IP scale and are used to evaluate hallux-related conditions.[4-7] Compared to other scales, the AOFAS hallux MTP-IP scale is more specific to hallux-related conditions and includes questions about the cosmetic appearance of the toe. Furthermore, it is based on both clinician-administered objective and patient-reported subjective questions, whereas some other outcome measures (e.g., MOXQF, SF-36, and FAAM) are based on patient-reported questions. However, all of these scales have their own strengths and weaknesses and may be used in different clinical situations depending on the specific needs of the patient and clinician.

In this study, we successfully translated and culturally adapted AOFAS hallux MTP-IP scoring scale into Turkish. Additionally, we evaluated the psychometric properties of the scale, including reliability and validity, in a group of patients with HV deformity. Our results suggest that the Turkish version of the AOFAS hallux MTP-IP has acceptable internal consistency, excellent test-retest reliability, good construct validity, and no floor/ceiling effects.[31] Taken together, the Turkish AOFAS hallux MTP-IP showed sufficient psychometric properties to assess pain, function, and alignment in patients with hallux pathologies in the Turkish population.[31]

Reliability and validation studies for AOFAS hallux MTP-IP were conducted for Italian,[14] Persian,[15] Arabic,[16] and Colombian Spanish.[17] In the present study, we used multi-stage guidelines for translation and cultural adaptation, as in all other versions. An acceptable level of internal consistency was obtained for the Turkish version of the AOFAS hallux MTP-IP, with Cronbach's $\alpha$ values of 0.704 and 0.706 for the total and functional scores, respectively. Compared to our results, Alhadhoud et al.[16] obtained higher internal consistency for the total score (Cronbach's $\alpha$=0.981) and function domain (Cronbach's $\alpha$=0.856). Molano Castro et al.[17] and Mahdaviazad et al.[15] reported similar internal consistencies in the Colombian Spanish (0.7) and Persian (0.72) versions, respectively.

Intra-rater reliability indicated excellent reliability for the overall score of the Turkish AOFAS hallux MTP-IP (0.971), as well as subscales of pain, function, and alignment (ICCs>0.9). In addition, a high inter-rater reliability was obtained for the total score (ICC=0.913). The Bland-Altman analysis illustrated perfect agreement between repeated measurements. Also, we obtained an acceptable random measurement error of MDC% was 19.8% for the total score. These results confirm that the Turkish AOFAS hallux MTP-IP is stable over time and provides consistent results for different examiners, given that the clinical condition of the patient is the same. Intra-rater reliabilities were reported as 0.86 and 0.811 in the Persian and Italian versions, respectively.[14,15] High ICCs were obtained in the Arabic (0.974)[16] and Persian (0.97)[15] versions for inter-rater reliability, whereas slightly lower

values were reported in the Italian version (0.881).[14] The reliability studies of AOFAS hallux MTP-IP in other languages did not provide agreement between measurements by Bland-Altman analysis and random measurement error. Therefore, we were unable to compare our results with those of other relevant studies.

Studies that discuss the construct validity of the AOFAS hallux MTP-IP investigated its relationship with many other outcome measures, including SF-12, SF-36, Visual Analog Scale (VAS), and FFI.[14-17] In these studies, the highest level of correlation was found to be with the physical summary score of SF-36 (r=0.50),[15] physical summary score of SF-12 (r=0.504),[16] general health component of SF-36 (r=0.584),[14] and pain domain of FFI (r=0.720).[17]

In the present study, we investigated the evidence for construct validity by assessing the correlations between the Turkish version of the AOFAS hallux MTP-IP and the Turkish versions of the SF-36[19] and MOXFQ.[18] We used SF-36, as it is among the most commonly reported outcome measures in validation studies of AOFAS scales. Additionally, we assessed construct validity using the MOXFQ, as it is widely used as an outcome measure of HV deformity. We found the highest correlations with the MOXFQ pain (r=-0.732) and walking/standing (r=-0.591) domains. As our study sample consisted of patients with HV, these results were expected. In addition, a moderate-to-good correlation was found with the physical component summary of SF-36 (r=0.485), similar to other validation studies in the literature. We observed weaker correlations with the social and mental health domains of the MOXFQ and SF-36. Our findings indicate good construct validity for the Turkish version of the AOFAS hallux MTP-IP. As no floor/ceiling effects was observed in the current study, we assumed good content validity. Similarly, no floor/ceiling effects have been reported in previous studies.

Nonetheless, this study has some limitations. First, the patients were selected using convenience sampling from a single center, which may have introduced a selection bias. Second, although the AOFAS hallux MTP-IP can be used to assess the clinical outcomes of patients with various hallux deformities, our study sample was limited to HV patients. Third, the number of patients evaluated for pre-testing was kept lower than the ideal number proposed by Beaton et al.,[21] as there were only three subjective items in the scale, and it was challenging to identify eligible patients. These issues may have affected the generalization of the results to different populations with hallux deformities in Türkiye, and further studies are needed to address these concerns to be more representative.

In conclusion, our study shows that the AOFAS hallux MTF-IP scale is translated and culturally adapted into Turkish with verified psychometric properties. Based on these results, the Turkish version of the scale can be used as a valid and reliable measure for clinical assessment of Turkish-speaking patients with HV deformities.

**Ethics Committee Approval:** The study protocol was approved by the Bolu Abant Izzet Baysal University Clinical Research Ethics Committee (date: 08.02.2022, no: 2022/14). The study was conducted in accordance with the principles of the Declaration of Helsinki.

**Patient Consent for Publication:** A written informed consent was obtained from each patient.

**Data Sharing Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Author Contributions:** Idea/concept, design, interpretation, writing the article: O.K., T.A.; Data collection: T.A., O.F.Y., M.T.T.; Analysis: O.K.; Literature review, final approval: O.K., T.A., O.F.Y., M.T.T.; Critical review, materials: O.K., T.A., O.F.Y., M.T.T.

**Conflict of Interest:** The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

**Funding:** The authors received no financial support for the research and/or authorship of this article.

## REFERENCES

1. Pamuk Ç. Metatarsal osteotomy for hallux valgus under wide-awake local anesthesia with no tourniquet technique. Jt Dis Relat Surg 2022;33:367-73. doi: 10.52312/jdrs.2022.664.
2. Roddy E, Zhang W, Doherty M. Prevalence and associations of hallux valgus in a primary care population. Arthritis Rheum 2008;59:857-62. doi: 10.1002/art.23709.
3. Churruca K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LED, et al. Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. Health Expect 2021;24:1015-24. doi: 10.1111/hex.13254.
4. Morley D, Jenkinson C, Doll H, Lavis G, Sharp R, Cooke P, et al. The Manchester-Oxford Foot Questionnaire (MOXFQ): Development and validation of a summary index score. Bone Joint Res 2013;2:66-9. doi: 10.1302/2046-3758.24.2000147.
5. Budiman-Mak E, Conrad KJ, Roach KE. The foot function index: A measure of foot pain and disability. J Clin Epidemiol 1991;44:561-70. doi: 10.1016/0895-4356(91)90220-4.
6. 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. doi: 10.1177/107110070502601113.
7. Martin RL, Burdett RG, Irrgang JJ. Development of the foot and ankle disability index (FADI). J Orthop Sports Phys Ther 1999;29:A32-3.

8. 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. doi: 10.1177/107110079401500701.
9. 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. doi: 10.1177/1938640019826680.
10. Ramada-Rodilla JM, Serra-Pujadas C, Delclós-Clanchet GL. Cross-cultural adaptation and health questionnaires validation: Revision and methodological recommendations. Salud Publica Mex 2013;55:57-66. doi: 10.1590/s0036-36342013000100009.
11. Şaylı U, Altunok EÇ, Güven M, Akman B, Biros J, Şaylı A. Prevalence estimation and familial tendency of common forefoot deformities in Turkey: A survey of 2662 adults. Acta Orthop Traumatol Turc 2018;52:167-73. doi: 10.1016/j.aott.2018.01.003.
12. Costello AB, Osborne J. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Pract Assess Res Eval 2005;10:1-9. doi: 10.7275/jyj1-4868.
13. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998;17:101-10. doi: 10.1002/(sici)1097-0258(19980115)17:1<101::aid-sim727>3.0.co;2-e.
14. Leigheb M, Vaiuso D, Rava E, Pogliacomi 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;90:118-26. doi: 10.23750/abm.v90i12-S.8978.
15. Mahdaviazad H, Kardeh B, Vosoughi AR. American Orthopedic Foot and Ankle Society hallux metatarsophalangeal-interphalangeal joint scale: A cross-cultural adaptation and validation study in the persian language. J Foot Ankle Surg 2020;59:729-32. doi: 10.1053/j.jfas.2020.01.006.
16. 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. doi: 10.1016/j.fas.2019.11.001.
17. Molano Castro JD, Sardoth Álvarez R, Franco Betancur A, Vargas Montenegro GE. Translation, cultural adaptation, and validation of the American Orthopedic Foot and Ankle Society scale in patients with hallux valgus in Colombia. J Foot Ankle Surg 2023;62:511-8. doi: 10.1053/j.jfas.2022.12.006.
18. Talu B, Bayramlar K, Bek N, Yakut Y. Validity and reliability of the Turkish version of the Manchester-Oxford foot questionnaire for hallux valgus deformity evaluation. Acta Orthop Traumatol Turc 2016;50:207-13. doi: 10.3944/AOTT.2015.15.0152.
19. Koçyiğit H, Aydemir O, Fişek G, Ölmez N, Memiş A. Kısa form-36 (SF-36)'nın Türkçe versiyonunun güvenilirliği ve geçerliliği.. İlaç ve Tedavi Dergisi 1999;12:102-6.
20. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
21. 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. doi: 10.1097/00007632-200012150-00014.
22. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2021.
23. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334. doi: 10.1007/BF02310555.
24. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979;86:420-8. doi: 10.1037//0033-2909.86.2.420.
25. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. Phys Ther 2006;86:735-43.
26. Smidt N, van der Windt DA, Assendelft WJ, Mourits AJ, Devillé WL, de Winter AF, et al. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. Arch Phys Med Rehabil 2002;83:1145-50. doi: 10.1053/apmr.2002.33728.
27. Heale R, Twycross A. Validity and reliability in quantitative studies. Evid Based Nurs 2015;18:66-7. doi: 10.1136/eb-2015-102129.
28. Ginty AT. Construct Validity. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York: Springer; 2013. p. 487-7.
29. Abma IL, Rovers M, van der Wees PJ. Appraising convergent validity of patient-reported outcome measures in systematic reviews: Constructing hypotheses and interpreting outcomes. BMC Res Notes 2016;9:226. doi: 10.1186/s13104-016-2034-2.
30. 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. doi: 10.1016/j.jclinepi.2006.03.012.
31. Atik OŞ. Which articles do the editors prefer to publish? Jt Dis Relat Surg 2022;33:1-2. doi: 10.52312/jdrs.2022.57903.

**APPENDIX-1**
Turkish version of AOFAS hallux MTP-IP Scale

**AOFAS Halluks Metatarsofalangeal-İnterfalangeal Ölçeği (Toplam 100 puan)**

| Maddeler | Puan |
|---|---|
| **Ağrı (40 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* | |
| Kısıtlılık yok | 10 |
| Çalışma hayatı gibi günlük yaşam aktivitelerinde kısıtlılık yok, eğlence aktivitelerinde kısıtlılık var | 7 |
| Günlük yaşam ve eğlence aktivitelerinde kısıtlılık var | 4 |
| Günlük yaşam ve eğlence aktivitelerinde ciddi kısıtlılık var | 0 |
| *Ayakkabı gereksinimleri* | |
| Modaya uygun, klasik ayakkabı, tabanlık ihtiyacı yok | 10 |
| Rahat ayakkabı, tabanlık var | 5 |
| Ortopedik ihtiyaçlar için modifiye edilmiş ayakkabı veya brace | 0 |
| *MTF eklem hareketi (dorsifleksiyon ve plantar fleksiyon toplamı)* | |
| Normal veya hafif kısıtlılık (75° veya daha fazla) | 10 |
| Orta derecede kısıtlılık (30°-74°) | 5 |
| Ciddi kısıtlılık (30°den az) | 0 |
| *IF eklem hareketi (plantar fleksiyon)* | |
| Kısıtlılık yok | 5 |
| Ciddi kısıtlılık (10° den az) | 0 |
| *MTF-IF stabilite (tüm yönlerde)* | |
| Stabil | 5 |
| Kesinlikle instabil veya yerinden çıkabilir | 0 |
| *Halluks MTF-IF ile ilişkili nasır* | |
| Nasır yok veya asemptomatik nasır var | 5 |
| Nasır var ve semptomatik | 0 |
| **Dizilim (15 puan)** | |
| İyi, halluks dizilimi iyi | 15 |
| Orta, halluks diziliminde hafif bozukluk var, semptom yok | 8 |
| Kötü, belirgin bir şekilde semptomatik dizilim bozukluğu var | 0 |

AOFAS: American Orthopaedic Foot and Ankle Society.