Foot and Ankle Surgery 28 (2022) 193–199

Contents lists available at ScienceDirect

Foot and Ankle Surgery

journal homepage: www.elsevier.com/locate/fas

Structural validity of the foot and ankle outcome score for orthopaedic pathologies with Rasch Measurement Theory

Krista Tapaninaho[a], Antti J. Saarinen[a], Outi Ilves[b], Mikko M. Uimonen[a], Arja H. Häkkinen[b,c], Henrik Sandelin[d,e,f], Jussi P. Repo[g,*]

[a] Department of Surgery, Central Finland Hospital District, Jyväskylä, Finland
[b] Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland
[c] Department of Physical Medicine, Central Finland Healthcare District, Jyväskylä, Finland
[d] Department of Orthopaedics and Traumatology, Vaasa Central Hospital, Vaasa and University of Helsinki, Helsinki, Finland
[e] Mehiläinen Sports Hospital, Vaasa, Finland
[f] Department of Orthopaedics and Traumatology, Helsinki University Hospital, Helsinki, Finland
[g] Department of Orthopaedics and Traumatology, Tampere University Hospital, Tampere, Finland

ARTICLE INFO

Article history:
Received 24 December 2020
Received in revised form 6 March 2021
Accepted 9 March 2021

Keywords:
Quality of life
Psychometric properties
Validity

ABSTRACT

Background: The Foot and Ankle Outcome Score (FAOS) is one of the most frequently used patient-reported outcome measures for foot and ankle conditions. The aim is to test the structural validity of the Finnish version of the FAOS using Rasch Measurement Theory.

Methods: FAOS scores were obtained from 218 consecutive patients who received operative treatment for foot and ankle conditions. The FAOS data were fitted into the Rasch model and person separation index (PSI) calculated.

Results: All the five subscales provided good coverage and targeting. Three subscales presented unidimensional structure. Thirty-eight of the 42 items had ordered response category thresholds. Three of the 42 items had differential item functioning towards gender. All subscales showed sufficient fit to the Rasch model. PSI ranged from 0.73 to 0.94 for the subscales.

Conclusions: The Finnish version of the FAOS shows acceptable structural validity for assessing complaints in orthopaedic foot and ankle patients.

1. Introduction

Patient-reported outcome measures (PROMs) are used to assess outcomes in both clinical practice and scientific research [1–3]. In foot and ankle surgery, a wide range of different PROMs are used [4,5], with over 140 different instruments reported in previous studies [3,6–10]. There is a wide variation in the psychometric properties of the foot and ankle PROMs [11]. An ideal PROM should be reliable, valid, and responsive [6,11].

The Foot and Ankle Outcome Score (FAOS) is an adaptation of the Knee Injury and Osteoarthritis Outcome Score [12]. FAOS is one of the most frequently reported PROMs in foot and ankle literature [11,13]. Prior studies of the FAOS have shown good measurement properties to assess various foot and ankle conditions [12,14–17]. FAOS has been validated in several languages [12,17–23].

Rasch analysis is a mathematical model which can be used for evaluation of psychometric properties of assessment instruments [30]. Rasch analysis is a combination of analyses involving calculation of the extent to which the observed responses fit to the pre-defined measurement model, assessment of unidimensionality of the scale and measurement precision [31,32]. The model is based on a theory of a latent trait and additive conjoint measurement [33]. It provides valuable information concerning the structural validity that cannot be obtained using other psychometric methods [34,35]. Although the FAOS has been extensively tested using the classical test theory [12,15–17,20,22,24–28] and other traditional psychometric tests [29], the psychometric properties have not been tested using the Rasch Measurement Theory.

The present analysis provides information on the suitability of FAOS for a generic sample of orthopedic foot and ankle conditions. The aim of the present study is to further test unidimensionality of the FAOS subscales, fit statistics, item residual correlation, coverage/targeting, differential item functioning (DIF) in gender, and person separation index using Rasch analytic techniques.

* Corresponding author.
E-mail address: mrjussirepo@gmail.com (J.P. Repo).

https://doi.org/10.1016/j.fas.2021.03.005
1268-7731/

2. Methods

Ethical approval was obtained from review board. The inclusion criteria were elective surgery for foot and/or ankle condition, age ≥18 years, and complete understanding of Finnish. The sample size adhered to guidelines for study design checklist for psychometric testing of PROMs [36].

2.1. Data collection

The data were collected between February 2018 and February 2019. Electronic data collection for the FAOS has previously been validated [37]. Patients completed a sociodemographic and clinical questionnaire charting their background data.

2.2. The Foot and Ankle Outcome Score (FAOS)

FAOS has a total of 42 questions divided into five subscales: pain (9 items), other symptoms (7 items), function in activities of daily living (17 items), function in sport and recreation (5 items), and quality of life (4 items). Each item is scored on a 5-point Likert-type scale. The item scores are summed and normalized into a subscale score from 0 (extreme symptoms) to 100 (no symptoms).

2.3. Translation and cross-cultural adaptation

The translation and cross-cultural adaptation were carried out according to the International Society of Pharmacoeconomics and Outcome Research guidelines [38].

Two native Finnish health care professionals fluent in English produced a forward translation of the original FAOS independently. Both translated versions were compared to each other, and possible discrepancies between the two independent translation versions were discussed. In this phase, only minor disagreements occurred. A consensus version was formed based on the discussion of two independent translators. In the next phase, a native English-speaker who has knowledge in medical terminology and skilled in Finnish language and culture translated the consensus version back into English. All three translators participated in reviewing the back-translation. There was discussion about the content of item S3 as there is no exact Finnish word for “catch” with similar meaning. The closest suitable Finnish term “*lukkiutua*” (“lock up”) was used with additional description “stop functioning suddenly”. Section title “Function, sports and recreational activities” was discussed as in Finnish language there was no need to separate the sports and recreational activities from each other. They were simply covered by term “*liikunta*”.

An expert panel reviewed the whole process and its different language versions and approved the pre-final version. The translation was pre-tested with patients. The expert panel reviewed the results and introduced the final version, which then underwent language editing by a language professional (available as supplementary material).

Table 1
Patients’ sociodemographic and clinical details.

	N = 218
Age (years), mean ± SD	55 ± 14
Female, *n* (%)	161 (74)
BMI[a] (kg/m^2), mean ± SD	28 ± 6
Duration of foot or ankle complaints (years), median (IQR[b])	8 (2–11)
Affected foot or ankle previously operated, *n* (%)	94 (43)
Indication for operative treatment, *n* (%)	
Deformity of foot or ankle	107 (49)
Osteoarthritis of foot or ankle	38 (17)
Flat foot or cavoid foot	8 (4)
Ankle instability	3 (1)
Other	62 (28)
General health state (Likert, 1–5), mean ± SD	3 ± 1
Education, *n* (%)	
Higher education	93 (43)
Upper secondary level education	28 (13)
Basic education	94 (43)

[a] Body mass index.
[b] Interquartile range.

2.4. Statistical methods

Data obtained from the patients was fitted to the Rasch model [39]. Description of the Rasch model and analysis process is explained in more detail elsewhere [30,32].

Unidimensionality is analyzed with the Rasch model to assess if the instrument measures a single trait [40]. The five subscales of the FAOS were tested separately. Principal component analysis (PCA) was used for defining the Rasch factor (first factor identified with the highest eigenvalue). Residual factors were divided into two groups according to the correlation coefficients (+0.3 and −0.3) with the second factor identified in the PCA. For each item, person estimates were calculated in both sets. The two item sets were compared using a series of independent samples t-tests. The authors hypothesized that each subscale would have less than 5% of significant t-tests indicating a unidimensional structure. In cases where unidimensionality was violated, testlets were produced [41]. These testlets are also referred to as super-items as they are formed from an item bundle. Item subsets were used with item residual loading of ±0.241. The probability for unidimensionality increases when items with high residual correlation are combined as it unifies the factor structure within the subscale. After testlet formation, another set of independent samples t-tests were conducted to investigate if violation of unidimensionality had been corrected. A threshold of 85% was used to reflect a unidimensional factor as total non-error variance.

Residual correlation refers to the extent to which two items of a scale have local dependency. In case of residual correlation, there is still independent correlation between items after the Rasch factor has been controlled. This reflects that the correlating items measure a different latent trait than the other items violating the assumption of unidimensionality. High residual correlation could potentially artificially reflect the reliability of a PROM. A value equal or above 0.2 was used to identify residual correlation. Residual correlation was hypothesized to be lower than this predefined threshold of 0.2.

Fit statistics were investigated for accuracy of data fit in the Rasch model. Item-person interaction (log residuals), item-trait interactions (chi-square [χ^2] values), and item characteristic curves were analyzed for item fit. Item and person fit residual between ±2.5 is generally considered acceptable. The authors hypothesized that non-significant *P*-values would be found after Bonferroni adjustment in χ^2-test, indicating a non-significant departure from the Rasch model.

Subscale coverage and targeting were investigated using a person and locations in a graphical threshold map. The distribution map curves, also called test information curves were illustrated to the maps using the Rumm. The test information curve is typically bell-shaped with its maximum at zero. The mathematics behind the information curve is explained in more detail in a text by Thomas Salzberger [42]. One-way analysis of variance was used to identify statistically significant difference between genders. The authors hypothesized that there would not be a difference between the genders. The statistical significance was set at 0.05.

Differential item functioning (DIF) was investigated to reveal whether there would be any difference between genders in answering the items of each subscale. Two types of DIF can be

Table 2
Item and person location and fit summary[a], person separation index and percentage of significant *t*-tests.

Subscale	Items		Persons		Chi-square	df	P	PSI (extrms/no exrtms)	Percentage (%) of significant *t*-tests
	Location (mean ± SD)	Fit residual (mean ± SD)	Location (mean ± SD)	Fit residual (mean ± SD)					
Pain	0.00 ± 1.83	−0.10 ± 0.95	−1.58 ± 2.20	−0.39 ± 1.28	31.0	18	0.03	0.83/0.88	7.6
Symptoms	0.00 ± 0.45	0.25 ± 1.76	−0.83 ± 1.04	−0.28 ± 1.11	59.1	14	<0.01	0.76/0.73	2.3
Function, daily living	0.00 ± 1.23	−0.09 ± 1.21	−3.28 ± 2.28	−0.36 ± 1.26	49.7	34	0.04	0.93/0.94	13.3
Testlet 1	0.00 ± 0.18	−1.10 ± 4.40	−3.27 ± 1.98	−0.46 ± 0.89	2.1	4	0.72	0.90/093	3.6
Sports	0.00 ± 0.95	−0.09 ± 1.75	−0.21 ± 2.12	−0.37 ± 1.06	15.3	10	0.13	0.87/0.84	4.9
HRQL	0.00 ± 0.75	0.24 ± 1.66	1.32 ± 1.88	−0.44 ± 1.10	8.7	8	0.37	0.81/0.77	4.2

[a] HRQOL: Health-related Quality of Life; DIF: differential item functioning; PSI: person separation index.

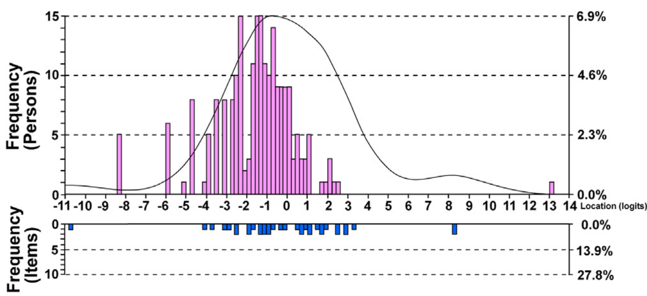

Fig. 1. Person-item threshold distribution map for the Pain subscale. Grouping set to Interval length of 0.20 making 125 groups. Total 217, mean −1.58, SD 2.20.

distinguished. In cases of uniform DIF, the difference in probability remains constant between different levels of measured trait in a single item. In non-uniform DIF, the groups have different probabilities at different levels of the measured trait. DIF for age was tested by dividing the age groups into two using mean age as a cut off. DIF for gender was tested dividing the study population to men and women. The authors hypothesized that there would be no DIF towards gender.

Person separation index (PSI) was calculated to examine sensitivity of the subscales to distinguish patients with varying state of foot condition. The PSI yields a value between 0 and 1, where lower and higher sensitivity is reflected. Lower PSI implies that the instrument may lack sensitivity to distinguish high and low performers. The authors hypothesized that the PSI would be at least 0.7. The threshold of 0.7 is relatively commonly used for indicating adequate PSI to distinguish between two groups.

Category probability curves were investigated to identify item thresholds between different response options. A threshold indicates a transition point between adjacent response categories, where the probability of responding in either response category becomes more likely in one category and less likely in the other category, with equal probability (50%) at the point of the threshold. Disordered threshold curves indicate that the response categories are not operating as they are supposed to. This could be due to various reasons including confusing response category wording, or an inappropriate number of distinct response categories, meaning that respondents have difficulties in distinguishing in which category their answer should fall into. The item response categories were hypothesized to form ordered thresholds.

3. Results

Altogether 267 patients agreed to participate in the present study, of which 49 were excluded due to incomplete data. Sufficient data was found for 218 patients for the Rasch analysis. Patients' sociodemographic details are provided in Table 1. Subscale fit statistics and unidimensionality values are presented in Table 2.

3.1. Pain subscale

A non-unidimensional structure was indicated (Table 2). Item reduction did not lead to a reasonable solution with unidimensionality. Therefore, testlets were created to conduct a subtest analysis based on residual correlations between items. No testlet solution was found. Residual correlation above 0.2 was found between 22 of the 36 item pairs (residual correlation matrix is presented in Appendices). No item misfit was found after Bonferroni adjustment in the Pain subscale.

The scale provided coverage for patients locating (Fig. 1). Location 13 (logits) was a patient with a hammer toe. Person-item distribution shows no difference between gender ($P = 0.8$). Item 5 "Walking on flat surface" had uniform DIF towards gender (Table 3). All items had ordered thresholds (Fig. 2).

3.2. Symptoms subscale

Symptoms subscale proved to be unidimensional (Table 2). Item 2 "Do you feel grinding, hear clicking or any other type of noise when your foot/ankle moves?" had a fit residual of 3.2 indicating potential item misfit. Residual correlation over 0.2 was found in 10 of the 36 item pairs (residual correlation matrix is presented in Appendices).

The scale provided coverage for patients (Fig. 3). Person-item distribution showed no difference between gender ($p = 0.5$). There was no DIF in any of the seven items towards gender (Table 3).

There were disordered thresholds in items 1, 4 and 5. Merging item response categories one, two and three in item 1, and response categories zero and one as well as three and four, respectively, led to ordered thresholds in each of the response categories (Fig. 4).

3.3. Function, daily living subscale

Item 12 "Lying in bed (turning over, maintaining foot/ankle position)" had fit residual of 2.9 indicating potential misfit (Fig. 5). Omitting the item 12 did not lead to a unidimensional structure as there were 12% of significant t-tests in equating item subsets. Residual correlation of 0.2 was found between 17 of the 136 item pairs. No clear testlet formation based on residual correlation and clinical relevance of items was available to achieve unidimensionality.

Item 15 had disordered thresholds. Merging the item response categories 2 and 3 created order in the thresholds (Fig. 6).

The scale provided coverage for patients locating (7.8). Person-item distribution showed no difference between gender ($p = 0.80$) or different health states ($P < 0.0001$). There was uniform DIF in

Table 3
Item specifications, fit statistics and differential[a] item functioning (DIF) analysis for gender.

Item	Location	Fit residual	DF	Chi-square	*P*	DIF gender (U/NON-U)
Pain						
1. How often do you experience f/a pain?	−4.042	1.882	177.10	4.256	0.12	–
2. Twisting/pivoting your f/a?	−0.735	−0.495	177.97	0.985	0.61	–
3. Straightening f/a fully?	0.243	−0.296	179.70	0.716	0.70	–
4. Bending f/a fully	0.076	0.235	177.97	0.262	0.88	–
5. Walking on flat surface	0.052	−0.441	181.44	4.079	0.13	U;P = 0.03
6. Going up or down stairs	−0.261	−0.556	180.57	4.712	0.09	–
7. At night while in bed	2.130	0.862	178.84	8.177	0.02	–
8. Sitting or lying	2.263	−0.972	178.84	5.134	0.08	–
9. Standing upright	0.275	−1.071	180.57	2.680	0.26	–
Symptoms						
1. Do you have swelling in your f/a?	−0.535	2.294	168.31	11.873	<0.01	–
2. Do you feel grinding, hear clicking or any other type of noise when your f/a moves?	−0.140	3.210	168.31	11.690	<0.01	–
3. Does your f/a catch or hang up when moving?	0.872	−0.316	168.31	3.855	0.15	–
4. Can you straighten your f/a fully?	−0.160	−1.194	168.31	6.869	0.03	–
5. Can you bend your f/a fully?	−0.266	−1.149	169.15	4.132	0.13	–
6. How severe is your f/a stiffness after sitting, lying or resting later in the day?	−0.015	−0.622	169.15	9.166	0.01	–
7. How severe is your foot/ankle stiffness after sitting, lying or resting later in the day?	0.242	−0.466	167.47	11.462	<0.01	–
Function, daily living						
1. Descending stairs	−1.335	0.570	179.45	1.941	0.38	–
2. Ascending stairs	−0.981	0.105	179.45	0.565	0.75	–
3. Rising from sitting	1.700	−1.646	179.45	0.546	0.76	–
4. Standing	−1.090	0.313	179.45	4.160	0.12	–
5. Bending to floor/pick up an object	−0.617	1.079	178.53	3.387	0.18	U; P = 0.01
6. Walking on flat surface	−0.770	1.118	179.45	1.365	0.51	–
7. Getting in/out of car	−0.351	−1.198	180.37	4.187	0.12	–
8. Going shopping	−0.858	0.003	178.53	0.729	0.69	–
9. Putting on socks/stockings	0.038	0.028	180.37	1.391	0.50	–
10. Rising from bed	0.752	0.450	180.37	0.625	0.73	–
11. Taking off socks/stockings	0.193	−0.278	180.37	0.182	0.91	–
12. Lying in bed (turning over, maintaining f/a position)	1.109	2.989	177.61	8.669	0.01	–
13. Getting in/out of bath	0.017	−1.029	173.93	3.345	0.19	–
14. Sitting	2.765	−1.292	175.77	2.355	0.31	–
15. Getting on/off toilet	0.090	−0.691	178.53	8.029	0.02	–
16. Heavy domestic duties	−2.008	0.055	177.61	1.363	0.51	–
17. Light domestic duties	1.348	−2.045	175.77	6.892	0.03	–
Sports						
1. Squatting	1.280	1.583	145.82	3.095	0.21	–
2. Running	−0.676	−0.390	140.39	2.445	0.29	–
3. Jumping	−0.920	−2.954	142.71	8.682	0.01	–
4. Twisting/pivoting on your injured f/a	−0.404	0.450	145.82	0.449	0.80	U;P = 0.003
5. Kneeling	0.720	0.856	144.27	0.588	0.75	–
HRQL						
1. How often are you aware of your f/a problem?	−1.047	2.077	146.88	1.397	0.50	–
2. Have you modified your lifestyle to avoid potentially damaging activities to your f/a?	0.718	1.142	147.62	0.417	0.81	–
3. How much are you troubled with lack of confidence in your f/a?	0.216	−0.751	146.88	1.297	0.52	–
4. In general, how much difficulty do you have with your f/a?	0.113	−1.523	147.62	5.566	0.06	–

[a] HRQOL: Health-related Quality of Life; DF: Degrees of freedom; DIF: differential item functioning; U/NON-U: Uniform / non-uniform.

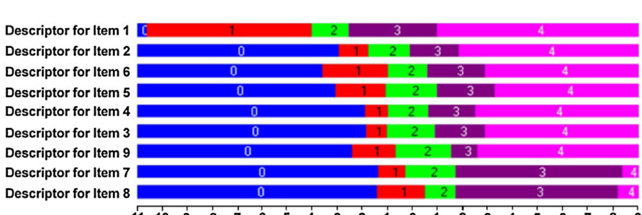

Fig. 2. Response category threshold map showing ordered thresholds in each of the items of the Pain subscale.

item 5 “Bending to floor/pick up an object” towards gender (Table 3).

3.4. Sports subscale

The Sports subscale had a unidimensional structure (Table 2). Residual correlation above 0.2 was found between eight of the nine item pairs. Item 3 “Jumping” had a fit residual of 3.0 indicating potential problems in item fit. All other items had sufficient item fit.

The scale provided coverage for patients located (Fig. 7). The sports subscale had no difference in coverage between gender ($p = 0.93$). There was uniform DIF towards gender in item 4 “Twisting/Twisting/pivoting on your injured foot or ankle” (Table 3). All items had ordered thresholds (Fig. 8).

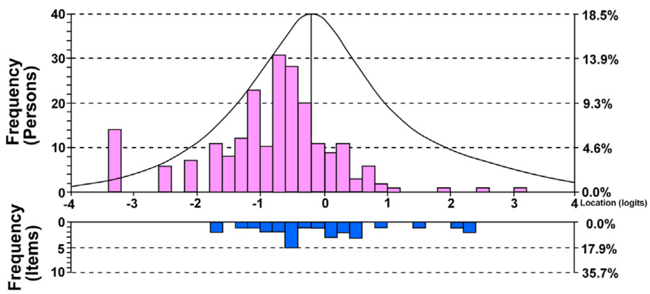

Fig. 3. Person-item distribution map for the Symptoms subscale. Grouping Set to Interval Length of 0.20 making 40 groups. Total 215, Mean −0.83, SD 1.04.

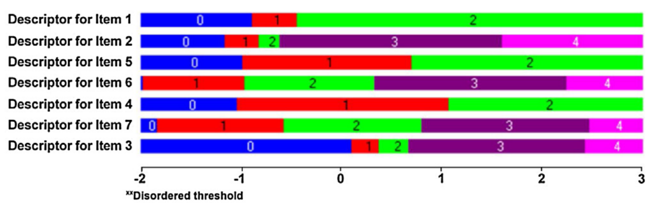

Fig. 4. Response category threshold map showing ordered thresholds after merging the response categories in items 1, 4 and 5 in the Symptoms subscale.

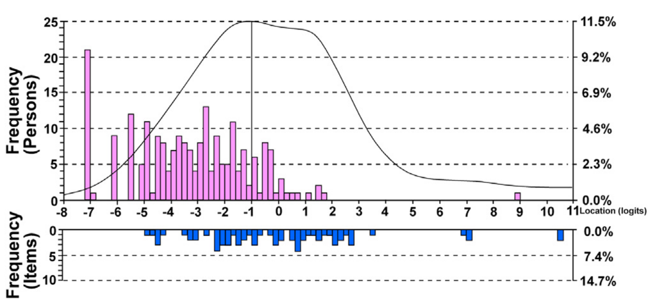

Fig. 5. Person-item distribution map for the Daily living subscale. Grouping set to Interval Length of 0.20 making 95 groups. Total 219, mean −3.28, SD 2.23.

3.5. Health-related quality of life subscale

The scale analysis showed a unidimensional structure (Table 2). All items had sufficient item fit. Residual correlation was found between five of the seven item pairs.

The scale provided coverage for patients (Fig. 9). There was no difference in the distribution between gender ($p = 0.76$). There was no DIF towards gender (Table 3). All four items had ordered thresholds (Fig. 10).

4. Discussion

Based on the results of the present study, the five subscales of the FAOS instrument provided good coverage and targeting for orthopaedic foot and ankle patients. Symptoms, Sports, and HRQoL subscales presented unidimensional structure.

The need to form testlets to obtain unidimensionality demonstrates that the items of the Function, Daily living, and Pain subscales might contain underlying constructs besides one latent trait. However, unidimensionality was achieved after testlet formation in the second analysis, demonstrating direct relation to the measured construct. The subscales can be considered to measure a unified set of complex experiences in people with foot and ankle problems. Although the unidimensionality of the FAOS subscales have not been previously studied using the Rasch model, several studies have conducted principal component or factor analysis for testing the unidimensionality [12,22,29]. Two prior studies found the subscales of the FAOS to be unidimensional [12,22]. One prior study reported unidimensionality in all subscales except the Symptoms subscale, which loaded on two factors in factor analysis [29].

All subscales provided good coverage and targeting for the patient sample. However, some patients exceeded the coverage of the instrument. There was no difference in coverage considering patient age or gender.

Item 12 in Function, daily living subscale “Lying in bed (turning over, maintaining foot/ankle position)” and item 3 in Sports subscale “Jumping” indicated potential misfits as they exceed the threshold for model fit. These items can both be considered opposite extremities in function; therefore, underlying conditions might explain these items having fit statistics exceeding the predefined thresholds. In the analysis, item 2 “Do you feel grinding, hear clicking or any other type of noise when your foot/ankle moves?” in the Symptoms subscale showed item fit residual above the predefined threshold. This finding might be due to the fact that

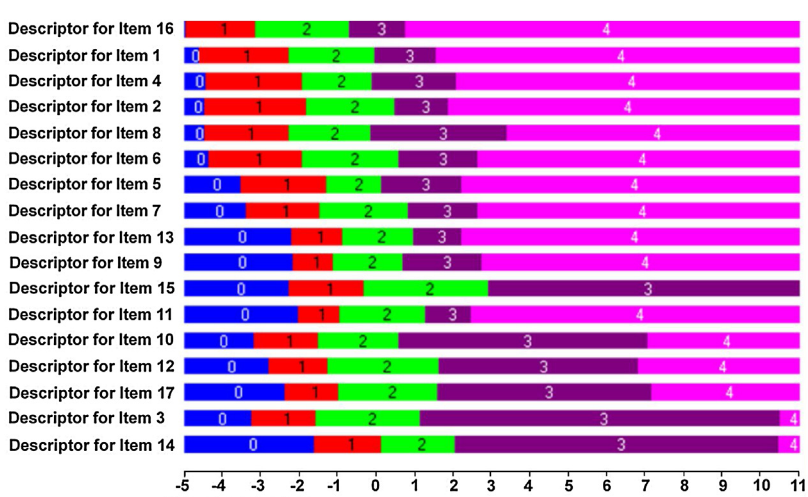

Fig. 6. Response category threshold map showing ordered thresholds after merging the response categories in item 15 in the Daily living subscale.

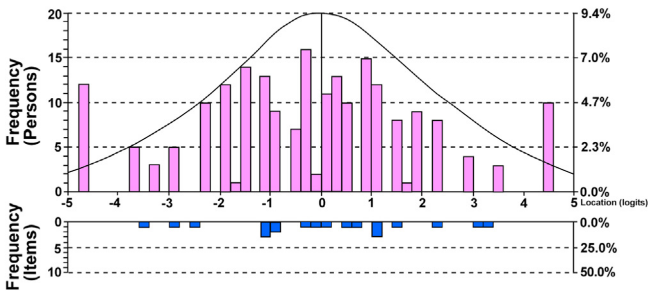

Fig. 7. Person-item distribution map for the Sports subscale. Grouping set to Interval length 0.20 making total 50 groups. Total 213, mean −0.21, SD 2.12.

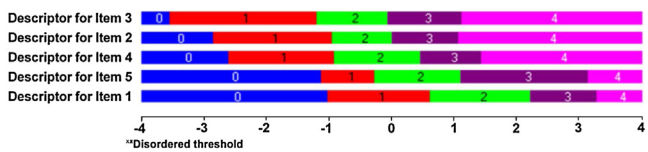

Fig. 8. Response category threshold map showing ordered thresholds all items in the Sports subscale.

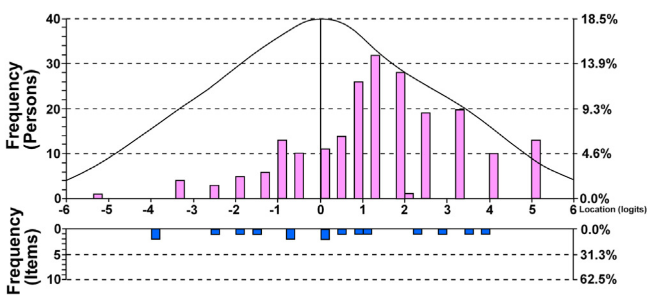

Fig. 9. Person-item distribution map for the HRQL subscale. Grouping set to Interval length 0.20 making total 50 groups. Total 216, Mean 1.321, SD 1.881.

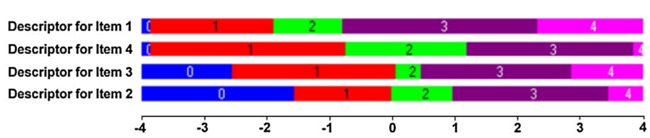

Fig. 10. Response category threshold map showing ordered thresholds in item 15 in the HRQL subscale.

most patients had arthrosis or deformity of the foot which are pathologies that may not cause irritating sounds. The item 2 is adapted from the KOOS instrument which is specific for knee problems [12]. This item might be more suitable for patients with knee problems or a specific group of foot and ankle patients that have symptoms causing grinding, clicking or other noises in movement.

According to person-item distribution and presence of DIF, the Symptoms and HRQL subscales showed good coverage and targeting as well as no DIF towards gender. The three other subscales had single item each with detectable DIF towards gender. However, there was no clear clinical indication why these items (‘Walking on flat surface’, ‘Bending to floor/pick up an object’, ‘Twisting/Twisting/pivoting on your injured foot or ankle’) would be different for men and women. The score was established for the Finnish version and it should be assessed further for cross-cultural DIF between Finnish and English versions for comparison.

Unidimensionality refers to the ability of measuring one specific trait. Unidimensional structure was found in three of the five subscales. In clinical practice, the subscales of the FAOS instrument can be used as separate entities to measure patient-reported outcomes or progress in rehabilitation, in respect of the scope of the assessment. Although the subscales of Pain and Function, daily living did not show a unidimensional structure, a clinician should not be discouraged in using the scales. In these two scales, there seems to be more than one underlying construct which is measured by the scale. All the subscales of the FAOS instrument have proven to be responsive to change [43]. Based on the present analyses, the FAOS seems to function well in patients who have undergone foot or ankle surgery. The FAOS could be implemented as a PROM in outpatient visits or clinical registries for foot and ankle patients.

The Rasch Measurement Theory provides a tool to investigate different aspects of the structure of a PROM. Thus, the study yields knowledge of the measurement properties of the FAOS instrument with a special emphasis on its structural validity, response category function, and robust analysis for coverage and targeting.

This study was limited by a generic sample of foot and ankle conditions which prevented us from examining specific pathologies. In future studies, FAOS can be analyzed for specific foot and ankle conditions using the Rasch method. A larger sample size for DIF analyses could have provided insight for DIF for different pathologies or co-morbidities.

In conclusion, a Finnish version of the FAOS was produced being equivalent in content and comprehensiveness to the original version. Despite minor violations of the Rasch model, the fit was acceptable. Combining or omitting some of the items could potentially lead to stronger construct validity in the subscales of Function, daily living and Pain. Nonetheless, the Rasch analysis provided evidence of acceptable structural validity for assessing complaints in orthopaedic foot and ankle patients.

Conflict of interest

None.

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.fas.2021.03.005.

References

[1] Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. BMJ 2015;350:g7818.
[2] Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract 2006;12 (5):559–68.
[3] Hunt KJ, Hurwit D. Use of patient-reported outcome measures in foot and ankle research. JBJS 2013;95(16):e118.
[4] Zwiers R, Weel H, Mallee WH, Kerkhoffs G, van Dijk CN, Ramos AJ, et al. Large variation in use of patient-reported outcome measures: a survey of 188 foot and ankle surgeons. Foot Ankle Surg 2018;24(3):246–51.
[5] Shazadeh Safavi P, Janney C, Jupiter D, Kunzler D, Bui R, Panchbhavi VK. A systematic review of the outcome evaluation tools for the foot and ankle. Foot Ankle Spec 2019;12(5):461–70.
[6] Button G, Pinney S. A meta-analysis of outcome rating scales in foot and ankle surgery: is there a valid, reliable, and responsive system? Foot Ankle Int 2004;25(8):521–5.
[7] Ponkilainen VT, Häkkinen AH, Uimonen MM, Tukiainen E, Sandelin H, Repo JP. Validation of the Western Ontario and McMaster Universities osteoarthritis index in patients having undergone ankle fracture surgery. J Foot Ankle Surg 2019;58(6):1100–7.

[8] Ponkilainen VT, Tukiainen EJ, Uimonen MM, Häkkinen AH, Repo JP. Assessment of the structural validity of three foot and ankle specific patient-reported outcome measures. Foot Ankle Surg 2020;26(2):169–74.
[9] Richter M, Agren P, Besse J, Cöster M, Kofoed H, Maffulli N, et al. EFAS Score – multilingual development and validation of a patient-reported outcome measure (PROM) by the score committee of the European Foot and Ankle Society (EFAS). Foot Ankle Surg 2018;24(3):185–204.
[10] Richter M, Agren P, Besse J, Coester M, Kofoed H, Maffulli N, et al. EFAS score-validation of Finnish and Turkish Versions by the Score Committee of the European Foot and Ankle Society (EFAS). Foot Ankle Surg 2020;26(3):250–3.
[11] Sierevelt IN, Zwiers R, Schats W, Haverkamp D, Terwee CB, Nolte PA, et al. Measurement properties of the most commonly used Foot-and Ankle-Specific Questionnaires: the FFI, FAOS and FAAM. A systematic review. Knee Surg Sports Traumatol Arthrosc 2018;26(7):2059–73.
[12] Roos EM, Brandsson S, Karlsson J. Validation of the foot and ankle outcome score for ankle ligament reconstruction. Foot Ankle Int 2001;22(10):788–94.
[13] Jia Y, Huang H, Gagnier JJ. A systematic review of measurement properties of patient-reported outcome measures for use in patients with foot or ankle diseases. Qual Life Res 2017;26(8):1969–2010.
[14] Mani SB, Brown HC, Nair P, Chen L, Do HT, Lyman S, et al. Validation of the Foot and Ankle Outcome Score in adult acquired flatfoot deformity. Foot Ankle Int 2013;34(8):1140–6.
[15] Hogan MV, Mani SB, Chan JY, Do H, Deland JT, Ellis SJ. Validation of the foot and ankle outcome score for hallux rigidus. HSS J 2016;12(1):44–50.
[16] Chen L, Lyman S, Do H, Karlsson J, Adam SP, Young E, et al. Validation of foot and ankle outcome score for hallux valgus. Foot Ankle Int 2012;33(12):1145–55.
[17] Angthong C. Validity and reliability of Thai version of the Foot and Ankle Outcome Score in patients with arthritis of the foot and ankle. Foot Ankle Surg 2016;22(4):224–8.
[18] Karatepe AG, Günaydın R, Kaya T, Karlıbaş U, Özbek G. Validation of the Turkish version of the foot and ankle outcome score. Rheumatol Int 2009;30(2):169–73.
[19] Larsen P, Boe AM, Iyer AB, Elsøe R. Danish translation of the Foot and Ankle Outcome Score. Dan Med J 2017;64(12).
[20] Lee KM, Chung CY, Kwon SS, Sung KH, Lee SY, Won SH, et al. Transcultural adaptation and testing psychometric properties of the Korean version of the Foot and Ankle Outcome Score (FAOS). Clin Rheumatol 2013;32(10):1443–50.
[21] Imoto AM, Peccin MS, Rodrigues R, Mizusaki JM. Translation, cultural adaptation and validation of Foot and Ankle Outcome Score (FAOS) questionnaire into Portuguese. Acta Ortop Bras 2009;17(4):232–5.
[22] Negahban H, Mazaheri M, Salavati M, Sohani SM, Askari M, Fanian H, et al. Reliability and validity of the foot and ankle outcome score: a validation study from Iran. Clin Rheumatol 2010;29(5):479–86.
[23] Pellegrini MJ, Poniachik R, Nuñez A, Escudero MI, Carcuro G, Cortes AA. Cross-cultural adaptation and validation of the Foot and Ankle Outcome Score (FAOS) into Spanish (Chile). Foot Ankle Surg 2020;26(7):790–6.
[24] Van Bergen C, Sierevelt IN, Hoogervorst P, Waizy H, Van Dijk CN, Becher C. Translation and validation of the German version of the foot and ankle outcome score. Arch Orthop Trauma Surg 2014;134(7):897–901.
[25] Ling SK, Chan V, Ho K, Ling F, Lui TH. Reliability and validity analysis of the open-source Chinese Foot and Ankle Outcome Score (FAOS). The Foot 2018;35:48–51.
[26] van den Akker-Scheek I, Seldentuis A, Reininga IH, Stevens M. Reliability and validity of the Dutch version of the Foot and Ankle Outcome Score (FAOS). BMC Musculoskelet Disord 2013;14(1):1–6.
[27] Sierevelt IN, Van Eekeren I, Haverkamp D, Reilingh ML, Terwee CB, Kerkhoffs G. Evaluation of the Dutch version of the Foot and Ankle Outcome Score (FAOS): responsiveness and minimally important change. Knee Surg Sports Traumatol Arthrosc 2016;24(4):1339–47.
[28] Mani SB, Do H, Vulcano E, Hogan MV, Lyman S, Deland JT, et al. Evaluation of the foot and ankle outcome score in patients with osteoarthritis of the ankle. Bone Joint J 2015;97(5):662–7.
[29] Golightly YM, Devellis RF, Nelson AE, Hannan MT, Lohmander LS, Renner JB, et al. Psychometric properties of the foot and ankle outcome score in a community-based study of adults with and without osteoarthritis. Arthritis Care Res (Hoboken) 2014;66(3 (Mar)):395–403.
[30] Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 2007;46(1):1–18.
[31] Andrich D. Rating scales and Rasch measurement. Expert Rev Pharmacoecon Outcomes Res 2011;11(5):571–85.
[32] Bond T, Yan Z, Heene M. Applying the Rasch model: fundamental measurement in the human sciences. Routledge; 2020.
[33] Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. Value Health 2004;7:S22–6.
[34] Ponkilainen VT, Miettinen M, Sandelin H, Lindahl J, Häkkinen AH, Toom A, et al. Structural validity of the Finnish Manchester-Oxford Foot Questionnaire (MOXFQ) using the Rasch model. Foot Ankle Surg 2021;27(1):93–100.
[35] Repo JP, Tukiainen EJ, Roine RP, Sampo M, Sandelin H, Häkkinen AH. Rasch analysis of the Lower Extremity Functional Scale for foot and ankle patients. Disabil Rehabil 2019;41(24):2965–71.
[36] 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.
[37] Uimonen MM, Ponkilainen VT, Toom A, Miettinen M, Häkkinen AH, Sandelin H, et al. Validity of five foot and ankle specific electronic patient-reported outcome (ePRO) instruments in patients undergoing elective orthopedic foot or ankle surgery. Foot Ankle Surg 2021;27(1):52–9.
[38] 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.
[39] Rasch, G. Probabilistic models for some intelligence and attainment tests. MESA Press, 5835 S. Kimbark Ave., Chicago, IL 60637; e-mail: MESA@uchicago.edu; web address: http://www.rasch.org; tele, 1993.
[40] Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 2007;S22–31.
[41] Wainer H, Kiely GL. Item clusters and computerized adaptive testing: a case for testlets. J Educ Meas 1987;24(3):185–201.
[42] Salzberger T. When gaps can be bridged. Rasch Meas Trans 2003;17(1)910–1 Available at: https://www.rasch.org/rmt/rmt171h.htm.
[43] Tapaninaho K, Uimonen MM, Saarinen AJ, Repo JP. Minimal important change for Foot and Ankle Outcome Score (FAOS). Foot Ankle Surg 2021 S1268-7731 (21)00019-9.