ARP Rheumatology 2025;4:123-137
https://doi.org/10.63032/YFBG1869

ORIGINAL ARTICLES

Reliability and validity of the European Portuguese version of the ABILHAND questionnaire in patients with systemic sclerosis

Albuquerque F[1], Neto M[1], Patrício R[2], Vieira M[3], Fernandes R[4], Silva A[1], Santiago T[1,5], Salvador MJ[1,5]

ABSTRACT

Systemic sclerosis (SSc) frequently affects hands, impairing its function and impacting quality of life. The ABILHAND-SSc is a patient-reported outcome measure (PROM) specifically designed to assess manual ability in SSc. Previously, an European Portuguese version has been translated and culturally adapted, but its psychometric properties remain to be validated. Therefore, we aim to validate the European Portuguese version of the ABILHAND-SSc in patients with systemic sclerosis (SSc) using Rasch analysis and classical test theory. A cross-sectional validation study was conducted in a tertiary rheumatology centre, including 53 Portuguese-speaking adults fulfilling ACR/EULAR 2013 classification criteria for SSc or VEDOSS classification criteria. Patients completed the ABILHAND-SSc along with other PROMs: HAQ-DI, SHAQ, SF-36v2, EQ-5D-5L, and ScleroID. Rasch analysis assessed item fit, dimensionality, targeting, and reliability. Test–retest reliability was evaluated in a stable subsample. Construct validity was examined via hypothesis testing and correlation with external instruments. The ABILHAND-SSc showed excellent internal consistency and Rasch-based reliability. Item fit statistics were within acceptable ranges, and no floor or ceiling effects were observed. Rasch and PCA analyses supported unidimensionality. Test–retest reliability was good. ABILHAND-SSc scores showed good correlation with related PROMs. Patients with diffuse cutaneous SSc had significantly lower manual ability than those with VEDOSS. In conclusion, the European Portuguese version of the ABILHAND-SSc is a valid, reliable, and feasible tool for assessing manual ability in patients with SSc. Its use in clinical practice and research may support patient-centred assessment and monitoring of hand function. Further validation in longitudinal and multicentre studies is warranted.

Keywords: Outcome measures; Validation; Manual ability; Physical function; Portuguese language; Questionnaire; Systemic sclerosis.

KEY MESSAGES

- First validation of the Portuguese ABILHAND-SSc assessing manual ability in systemic sclerosis
- Further validation in other Portuguese cohorts is needed to ensure the generalizability of these findings

INTRODUCTION

Systemic sclerosis (SSc) is a chronic, progressive autoimmune disease characterized by fibrosis, vasculopathy and immune system abnormalities[1]. With no known cure, SSc significantly impacts patients' quality of life due to its multisystem involvement and diverse clinical manifestations[2]. In this context, patient-reported outcome measures (PROMs) have emerged as critical tools for capturing the patient's perspective and addressing the complex needs associated with this disease[3]. PROMs play a vital role in monitoring disease impact and guiding interventions aimed at improving the quality of life of individuals living with SSc[3].

The hand is frequently affected in SSc, with fibrotic changes in the skin and vascular abnormalities contributing to Raynaud's phenomenon, which is often complicated by digital ulcers and, in severe cases, gangrene[4]. Hand involvement, especially hand contracture and tendon friction rubs, may lead to func-

[1] Serviço de Reumatologia, Unidade Local de Saúde de Coimbra, Portugal; [2] Centro de Medicina de Reabilitação da Região Centro-Rovisco Pais, Unidade Local de Saúde de Coimbra, Portugal [3] Serviço de Oncologia Médica, Instituto Português de Oncologia, Portugal; [4] Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, United Kingdom; [5] Faculdade de Medicina, Universidade de Coimbra, Portugal

Submitted: 13/04/2025
Accepted: 07/06/2025

Correspondence to: Fernando Albuquerque
E-mail: fernando.mna@gmail.com

tion impairment and, consequently, the performance of daily activities[5]. Despite the importance of assessing hand function, many PROMs used in this context have been adapted from tools designed for other conditions, which presents challenges in adequately capturing the unique range of impairments in SSc[2].

The ABILHAND questionnaire was originally developed to specifically assess manual ability in patients with rheumatoid arthritis[6] and has since been validated for use in individuals with systemic sclerosis[7]. ABILHAND is a patient-centred PROM that evaluates the perceived difficulty of performing daily manual activities, offering a comprehensive assessment of hand function across various contexts. It includes items that assess different aspects of manual function, scored based on the perceived difficulty of execution, providing a reliable measure of functional limitations[6, 7].

The ABILHAND questionnaire has been translated and cross-culturally adapted into European Portuguese following a forward-backward translation method, followed by a review by Portuguese healthcare professionals specialized in Systemic Sclerosis, and a field test with cognitive debriefing by SSc patients. Face validity of the original ABILHAND has been ensured by the involvement of patients and experts in all steps of its development, including item selection, question formulation and scoring methods. Similarly, the face validity of the European Portuguese translation was ensured by the involvement of healthcare professionals and patients[8].

METHODS

Design, setting and study population

This study was conducted using a cross-sectional design. Participants were consenting adults consecutively selected from all patients receiving care in the rheumatology department of a tertiary hospital who fulfilled the 2013 American College of Rheumatology (ACR)/ European Alliance of Associations for Rheumatology (EULAR) criteria for the classification of SSc (9) or the Very Early Systemic Sclerosis (VEDOSS) criteria of the European Scleroderma Trials and Research Group (EUSTAR) (10). Exclusion criteria were not being a native Portuguese speaker and reader and not being able to provide consent. Data was collected between January 2022 and June 2024.

Data collection and variable definitions

One set of data was collected from each patient at one clinical visit, including age, sex, years of education, and employment status. Also, disease subtype (limited cutaneous SSc [lcSSc]; diffuse cutaneous SSc [dcSSc]; SSc sine scleroderma, VEDOSS), disease duration, clinical manifestations, and positivity for antinuclear antibodies (ANAs) and specific SSc antibodies.

Additionally, the following measurement instruments were collected, by self-completion of the Portuguese versions of the questionnaires:

- The ABILHAND-SSc, a measure of hand ability validated in SSc patients[7], comprised of 26 items, which are manual tasks rated by each patient according to their ability to perform them (impossible, difficult or easy). The final score is a measure of the patient's hand ability, that ranges from 0 to 100%. The score has been subjected to a Portuguese translation and cross-cultural adaptation, but this version is yet to be validated[8]. The Portuguese version of the ABILHAND-SSc questionnaire can be found in the supplemental materials (Appendix I).

- Short Form Health Survey (SF-36v2), a patient-reported measure of functional health and well-being comprised of 8 dimensions: physical functioning (PF), bodily pain (BP), role limitations due to physical health (RP), general health perception (GH), mental health (MH), role limitations due to emotional problems (RE), vitality—(VT), and social functioning (SF). The final score ranges from 0 (death) to 100 (perfect health status). The questionnaire has been translated into European Portuguese, and the Portuguese population norms have been established[11].
- European Quality of Life 5 Dimensions 5 Levels (EQ-5D-5L), a self-reported measure of health-related quality of life for clinical and economic appraisal. It is composed of five dimensions, scored from 1 (best state) to 5 (worst state): mobility, self-care, usual activities, pain/discomfort, anxiety/depression. Each health state is assigned a unique score using a value set derived from specific populations. The value set for EQ-5D-5L has been determined for the Portuguese population, and the index ranges from -0.603 (worst state) to 1 (best state)[12]. Additionally, a visual analogue scale (EQ-VAS) is scored, ranging from 0 (worst imaginable health) to 100 (best imaginable health).
- Scleroderma Health Assessment Questionnaire (SHAQ), comprised of the Health Assessment Questionnaire Disability Index (HAQ-DI) and six additional visual analogue scales (VASs) – pain, GI symptoms, breathing, Raynaud's phenomenon, finger ulcer, and overall disease severity. The HAQ-DI contains 20 items and measures eight domains: dressing and grooming, arising, eating, walking, hygiene, reach, grip, and activities[13]. Each item is scored from 0 (without difficulty) to 3 (unable to do). The highest-scored item in each domain determines the total score for that domain, except for the necessity of

aids or devices, in which case the minimum score for the domain is 2. The total score is the average of the domains and ranges from 0 to 3. Each additional VAS has a 1-week recall period and ranges from 0 to 100mm. Recently, the Portuguese version of the SHAQ was validated[14].

- ScleroID, a disease-specific patient-reported measure of SSc disease burden. The questionnaire is comprised of 10 items, scored from 0 (no impact) to 10 (extreme impact): Raynaud's, hand function, upper GI symptoms, pain, fatigue, lower GI symptoms, life choices and activity limitation, body mobility, dyspnea, and digital ulcers. Each item is multiplied by a weight and the total score ranges from 0 to 10. The face validity, construct validity, internal consistency, test-retest reliability and sensitivity to change of the English version were previously tested and considered satisfactory[15]. The translation and cross-cultural adaptation of the original English version into European Portuguese was recently accomplished[16], but this version of the questionnaire is yet to be validated in Portuguese patients.

ABILHAND-SSc validation

Feasibility and missing data

Feasibility was assessed through item response rate, with items displaying response rates below 80% considered potentially irrelevant. Missing responses in the ABILHAND-SSc were handled using the Rasch model for imputation, which estimates the most probable response based on the individual's ability and the item difficulty.

Rasch analysis

Rasch analysis was conducted using the partial credit model (PCM). The model estimates person ability on a linear logit scale and evaluates the internal structure of the ABILHAND-SSc through several psychometric indicators. Item–person targeting was assessed by comparing the average person ability with the average item difficulty, which is fixed at zero logits. Adequate targeting was considered when person and item distributions aligned on the logit scale. Item fit was evaluated using infit mean square (MnSq) statistics and standardized residuals (t-statistics), with misfitting items defined as those with residuals exceeding ±2.5 or statistically significant Bonferroni-adjusted chi-square values ($p < 0.002$). Scale reliability was assessed via the Person Separation Index (PSI), which reflects the instrument's ability to distinguish among different levels of manual ability. Higher values indicate better discrimination, with PSI ≥ 0.90 considered sufficient to differentiate at least four distinct ability strata. To explore the dimensionality of the construct, principal component analysis (PCA) of the residuals was conducted, and unidimensionality assumption was considered met when the first factor explained a substantial proportion of the variance. In addition, Differential Item Functioning (DIF) was examined through two-way ANOVA on item residuals across strata of person ability and patient factors, including sex, age group, disease duration, SSc subtype and skin thickening proximal to metacarpophalangeal joints. DIF occurs when individuals from different groups with the same underlying level of manual ability have a different probability of endorsing an item, indicating potential item bias. A Bonferroni-adjusted significance level of $p < 0.002$ was used to identify meaningful DIF. A person–item map was generated to visually inspect the distribution of item difficulties in relation to person abilities.

Reliability

Internal consistency was considered acceptable if Cronbach's alpha ≥ 0.7. Rasch-based reliability was additionally assessed through the PSI, as referred above. Test–retest reliability was assessed using the intraclass correlation coefficient (ICC). Patients in a stable disease state, defined as having no changes in disease state or treatment according to clinician judgment, were invited to retake the ABILHAND-SSc within 15 to 30 days after their initial response and to return the completed questionnaire by mail. ICC estimates and 95% confidence intervals were calculated based on a single-rating, absolute-agreement, 2-way mixed-effects model. ICCs were interpreted as follows: <0.5 – poor; 0.5 ≤ ICC ≤ 0.75 – moderate; 0.75 < ICC ≤ 0.90 – good; >0.90 – excellent.

Construct Validity

Construct validity was examined through multiple complementary strategies. Unidimensionality of the scale was assessed using principal component analysis (PCA) of the residuals within the Rasch model. Known-groups validity was evaluated by comparing ABILHAND-SSc scores across predefined subgroups according to sex, age, disease duration, and SSc subtype. Convergent validity was assessed through Spearman's or Pearson's correlation between ABILHAND-SSc scores and external instruments measuring related constructs: HAQ-DI and SHAQ (disability), SF-36 (functional health and well-being), EQ-5D (quality of life), and ScleroID (disease burden). Floor and ceiling effects were considered present if more than 15% of patients obtained either the minimum or maximum possible score.

Statistical analysis

Continuous variables were described as means and standard deviations (SD) or medians and interquartile ranges (IQR), according to the normality of their

distribution. Normality was assessed by computing z-scores for skewness and kurtosis (normality assumed if z-score was within ±2.58) for samples ≥50, and by Shapiro-Wilk test for samples <50. Categorical variables were described as proportions (%). For continuous variables, between-group comparisons were made using the t-test or Mann-Whitney U test (dichotomous independent variable), and one-way ANOVA or Kruskal-Wallis H tests (categorical variable with >2 groups), depending on the distribution and presence of outliers. Associations between continuous variables were assessed using Pearson's or Spearman's correlation coefficients, based on the normality and linearity of relationships. Correlation strength was classified as: low (≤0.3), moderate (0.3–0.49), good (0.5–0.79), or very good (≥0.8). Statistical significance was set at $p < 0.05$. Analyses were performed using IBM SPSS Statistics (v. 26.0.0.0) and The R Project for Statistical Computing (v. 4.4.3).

RESULTS

Sample

A total of 53 patients were included in this study, with 84.8% of them being female. The mean age was 58.7 years and the median duration of the disease was 10.9 years. The majority of patients had lcSSc (58.5%), followed by 28.3% with dcSSc, 5 patients were classified as VEDOSS and 2 with SSc sine scleroderma. Sociodemographic characteristics, lifestyle factors, and disease-related information are detailed in Table I. Table II presents the results from the EQ-5D-5L, SF-36v2, SHAQ and ScleroID measurement tools. The mean ABILHAND-SSc total measurement was 37.23.

Feasibility and missing data

Feasibility was assessed by evaluating item response rates and patterns of missing data. Of the 26 items included in the ABILHAND-SSc, the item "Winding up a wristwatch" had 9.4% of missing responses, and five items had 1.9% of missing responses, indicating good acceptability of the scale among patients (Table III). Missing responses were handled using the Rasch model.

Rasch analysis

Rasch analysis supported the internal structure of the scale. The mean person ability was –0.06 logits, while the mean item difficulty was –2.38 logits, resulting in a targeting difference of 2.32 logits. This suggests that most patients had greater manual ability than the average difficulty of the items. The distribution of patient abilities and item thresholds is shown in Appendix II and III, respectively. The person–item map (Figure 1) visually confirms this targeting imbalance, with most item thresholds clustering at lower difficulty levels compared to the distribution of patient abilities.

Item fit statistics indicated that all items conformed to the Rasch model expectations, with infit t-statistics within the acceptable range (–2.5 to +2.5) and non-significant Bonferroni-adjusted χ^2 values (Appendix IV), supporting good compatibility of the items with the unidimensional Rasch construct. The Rasch PSI was 0.951, indicating excellent reliability and suggesting that the ABILHAND-SSc was able to discriminate at least four distinct levels of manual ability within the sample.

In PCA of residuals, the proportion of variance explained by the first factor was 52.9%, supporting a dominant latent trait and thereby supporting the unidimensionality assumption of the Rasch model.

Differential Item Functioning (DIF) analysis identified no significant DIF by sex or disease duration. However, three items displayed DIF by age group, four by presence of proximal skin thickening, and six by SSc subtype (Appendix V).

Reliability

Internal consistency of the ABILHAND-SSc was excellent, as indicated by a Cronbach's alpha of 0.966 based on the 26 items included in the scale. Item–total correlations were all positive and ranged from 0.462 to 0.850, suggesting that all items contributed meaningfully to the overall construct (Appendix VI). No substantial improvement in alpha was observed upon item deletion, supporting the coherence and homogeneity of the scale items. Rasch-based reliability was also high, as reflected by a PSI of 0.951.

Although most patients were in a stable disease state, adherence to the retest was low, with 13 patients (24.5% of the whole cohort) mailing back the results, all of which had lcSSc (38.5%) or dcSSc (61.5%). The ICC was 0.795 (95% CI: 0.471 to 0.932; $p < 0.001$), indicating good reliability over time and supports the temporal stability of the instrument.

Construct Validity

Construct validity was examined through multiple complementary strategies. Unidimensionality of the scale was assessed through principal component analysis of the Rasch residuals, as described above.

Regarding floor and ceiling effects, one patient (1.9%) obtained the minimum score and five patients (9.4%) achieved the maximum score. As both frequencies were below the predefined 15% threshold, no floor or ceiling effects were observed.

Known-groups validity was supported by significant differences in ABILHAND-SSc scores between certain

TABLE I. Demographic and disease characteristics.

		Missingness (%)
Age (years) – mean ± SD	58.7 ± 12.6	0
Female sex – n (%)	45 (84.9)	0
SSc subtype		0
lcSSc – n (%)	31 (58.5)	
dcSSc – n (%)	15 (28.3)	
Sine scleroderma – n (%)	2 (3.8)	
VEDOSS – n (%)	5 (9.4)	
Disease duration – median (IQR)	10.9 (12.3)	0
Clinical manifestations		
Skin thickening proximal to MCPs – n (%)	22 (41.5)	0
Puffy fingers – n (%)	17 (32.1)	0
Sclerodactyly – n (%)	32 (60.4)	0
Digital ulcers – n (%)	22 (41.5)	0
Pitting scars – n (%)	18 (34.0)	0
Telangiectasia – n (%)	35 (66.0)	0
Capillaroscopy abnormalities – n (%)	35 (66.0)	0
Pulmonary arterial hypertension – n (%)	4 (7.5)	0
Interstitial lung disease – n (%)	11 (20.8)	0
Raynaud's – n (%)	51 (96.2)	0
Arthralgia – n (%)	28 (52.8)	0
Myositis – n (%)	4 (7.5)	0
Upper GI involvement – n (%)	33 (62.3)	0
Lower GI involvement – n (%)	7 (13.2)	0
Renal involvement – n (%)	1 (1.9)	0
Immunological profile		
ANAs – n (%)	52 (98.1)	0
Anti-centromere – n (%)	28 (52.8)	0
Anti-topoisomerase I – n (%)	15 (28.3)	0
Employment status		1.9
Employed – n (%)	18 (34.6)	
Unemployed – n (%)	4 (7.6)	
Retired – n (%)	30 (57.7)	
Education		1.9
No formal education – n (%)	1 (1.9)	
4 years – n (%)	14 (26.9)	
5-12 years – n (%)	24 (46.2)	
>12 years – n (%)	13 (25.0)	

ANAs: antinuclear antibodies; dcSSc: diffuse cutaneous systemic sclerosis; GI: gastrointestinal; lcSSc: limited cutaneous systemic sclerosis; MCPs: metacarpophalangeal joints; SD: standard deviation; VEDOSS: very early diagnosis of systemic sclerosis

TABLE II. Patient reported outcomes

		Min-Max	Mean±SD	Median (IQR)	Missingness (%)
EQ-5D-5L	Index	0.19-1.00	0.71±0.21	—	0
	VAS	4-90	60.5±20.4	—	1.9
SF-36v2	Physical functioning	5-100	52.5±24.0	—	0
	Physical role functioning	0-100	43.2±25.6	—	1.9
	Bodily pain	0-100	43.7±22.1	—	0
	General health perceptions	0-75	33.4±16.2	—	0
	Vitality	0-87.5	36.2±20.7	—	0
	Social role functioning	0-100	61.1±24.7	—	0
	Emotional role functioning	0-100	51.2±25.1	—	1.9
	Mental health	10-100	55.0±24.4	—	0
SHAQ	HAQ-DI	0-2.22	0.99±0.63	—	0
	Pain VAS	1-86	42.4±24.5	—	1.9
	GI symptoms VAS	0-100	27.7±30.1	—	1.9
	Breathing VAS	0-86	25.1±28.7	—	1.9
	Raynaud's VAS	0-88	41.6±29.7	—	1.9
	Finger ulcer VAS	0-92	25.2±30.2	—	1.9
	Overall disease severity VAS	0-90	47.0±27.5	—	1.9
UCLA-GIT 2.0	Reflux	0-2.13	0.61±0.57	—	5.7
	Distension/bloating	0-3	1.01±0.85	—	1.9
	Faecal soilage	0-3	—	0.00 (0.00)	1.9
	Diarrhoea	0-2	—	0.50 (1.00)	1.9
	Social functioning	0-2.67	—	0.00 (0.54)	5.67
	Emotional wellbeing	0-2.89	—	0.22 (0.86)	9.4
	Constipation	0-2.50	—	0.50 (0.75)	7.6
	Total score	0-2.02	—	0.36 (0.60)	13.2
ScleroID	Raynaud's	0-10	5.02±2.87	—	1.9
	Hand Function	0-10	5.85±2.82	—	0
	Superior GI tract	0-9	3.72±3.11	—	0
	Pain	0-10	5.38±3.01	—	0
	Fatigue	0-10	5.79±2.92	—	0
	Lower GI tract	0-9	4.29±3.24	—	1.9
	Daily activities	0-10	4.92±2.80	—	1.9
	Mobility	0-10	4.75±3.26	—	0
	Dyspnoea	0-9	2.67±2.90	—	0
	Ulcers	0-9	2.68±3.26	—	0
	Total score	0,33-8,40	4.60±2.37	—	0

GI: Gastrointestinal; IQR: interquartile range; Max: maximum; Min: minimum; SD: standard deviation; VAS: Visual analogue scale

TABLE III. Proportion of missing responses per item

Item	Missing (%)
Threading a needle	0.0
Putting on a piece of jewellery	1.89
Cutting meat	0.0
Handling scissors	0.0
Taking the cap off a bottle	0.0
Taking a coin out of a pocket	0.0
Cutting one's nails	0.0
Unwrapping a chocolate bar	0.0
Wiping windows	0.0
Turning off a tap	0.0
Lacing shoes	0.0
Handling a stapler	0.0
Opening mail	0.0
Peeling onions	1.89
Winding up a wristwatch	9.43
Peeling potatoes with a knife	0.0
Opening a screw-topped jar	0.0
Fastening the zipper of a jacket	1.89
Cleaning vegetables	0.0
Spreading butter on a slice of bread	0.0
Putting cream on one's body	1.89
Tearing open a package of chips	0.0
Brushing one's hair	1.89
Fastening a snap-fastener (e.g., bag, jacket)	0.0
Shelling hazel nuts	0.0
Buttoning up trousers	0.0

Percentage of missing responses per ABILHAND-SSc item. Only one item ("Winding up a wristwatch") had a missing rate above 5% (9.4%), while all others were below 2%, suggesting good feasibility and item comprehensibility.

clinical subgroups (Table IV). While no significant differences were observed by sex, age group or disease duration, patients with diffuse SSc presented lower scores than those with VEDOSS ($p = 0.034$), and patients without proximal skin thickening scored significantly higher than those with this clinical sign ($p = 0.044$).

Convergent validity was confirmed by correlation analysis between the ABILHAND-SSc and other patient-reported outcomes (Table V). Strong correlations were observed with the EQ-5D-5L index, SF-36 physical functioning, SF-36 general health perceptions, HAQ-DI, SHAQ overall, and ScleroID scores. Moderate correlations were found with the remaining measures, except for the SHAQ finger ulcer VAS, which showed a low correlation. All associations were in the expected direction, supporting the theoretical consistency of the ABILHAND-SSc with related constructs.

DISCUSSION

This study aimed to validate the European Portuguese version of the ABILHAND-SSc questionnaire in a cohort of patients with systemic sclerosis. Our findings support the validity, reliability, and clinical utility of the instrument to assess manual ability in Portuguese-speaking individuals with SSc.

The questionnaire demonstrated excellent internal consistency, strong test–retest reliability, and a high Person Separation Index, suggesting that the ABILHAND-SSc is capable of reliably distinguishing between different levels of manual ability. Rasch analysis confirmed good item–person targeting, although the mean person ability exceeded the mean item difficulty, indicating that most patients had greater manual ability than the average item challenge. Despite this, all items displayed good fit statistics, supporting the unidimensional structure of the scale and its coherence as a single construct.

Construct validity was supported by significant correlations with other validated patient-reported outcomes. The ABILHAND-SSc showed good correlation with the SF-36 physical functioning and general health domains, HAQ-DI, SHAQ, EQ-5D-5L index, and ScleroID, aligning with its theoretical framework as a measure of upper limb functional limitation. The low correlation with the SHAQ finger ulcer VAS may reflect the multifactorial nature of ulcer-related disability, which may not always directly impair performance of the daily tasks covered in the ABILHAND-SSc.

The absence of floor and ceiling effects further supports the capacity of the ABILHAND-SSc to measure a broad range of manual ability levels. The PCA of residuals and factor analysis both confirmed the unidimensionality of the scale.

Significant differences were found between disease subtypes, with patients with diffuse SSc reporting worse manual ability than those with VEDOSS, in accordance with the known association between diffuse skin involvement and greater functional disability. The lack of significant sex- or age-related differences supports the scale's generalizability, although some items exhibited

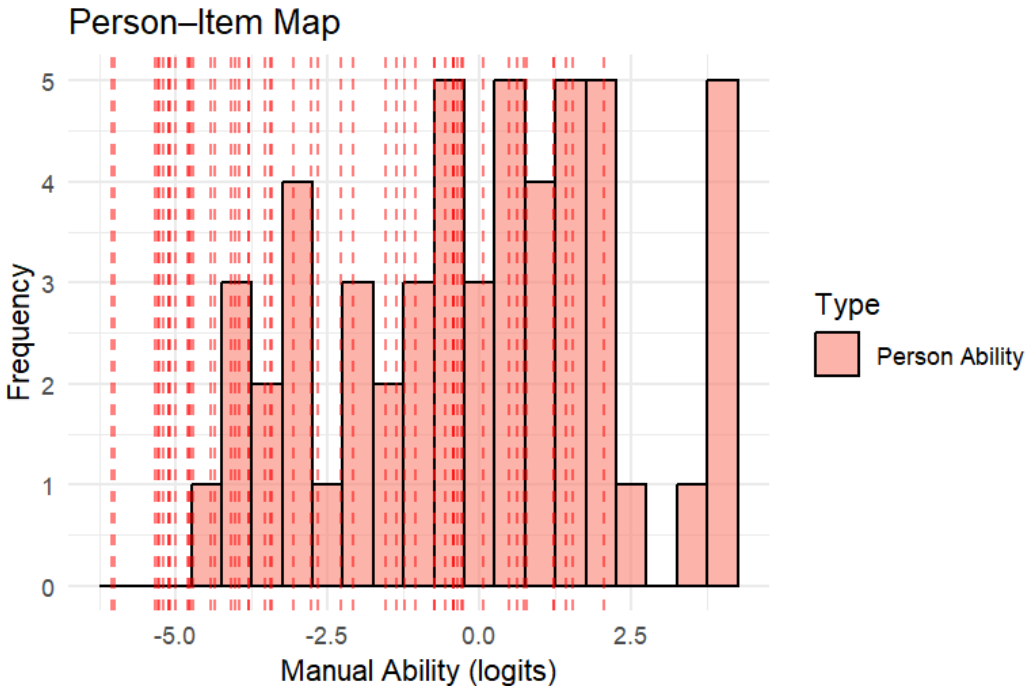


Figure 1. Person–item map showing the distribution of person abilities (in pink bars) and item thresholds (red dashed lines) on the same logit scale. Each item contributes two thresholds, corresponding to the transition points between response categories.

TABLE IV. ABILHAND-SSc score stratified by sociodemographic and clinical variables

		N	Mean ± SD	Median (IQR)	Statistic	p-value
Sex	Female	45	48.8 ± 28.1	—	t(51) = -0,79	0.434
	Male	8	57.2 ± 25.9	—		
SSc subtype*#	Diffuse	15	—	42.9 (27.6)	H(3) = 9.46	0.024
	Limited	31	—	54.5 (47.5)		
	VEDOSS	5	—	71.3 (21.1)		
	Sine	2	—	81.1 (NA)		
Age*+	<50 years	12	—	54.8 (26.0)	H(2) = 4.04	0.133
	[50-70[years	30	—	50.1 (38.0)		
	≥70 years	11	—	13.9 (53.5)		
Disease duration	<5 years	13		49.8 (26.7)	U = 253.50, z = -0.13	0.893
	≥5 years	40		56.3 (42.8)		
Skin thickening (proximal to MCPs)	No	22	59.4 ± 27.8		t(51) = 2.14	0.037
	Yes	31	43.4 ± 26.0			

MCPs: metacarpophalangeal joints; NA: not applicable; QR: interquartile range; H: Kruskal-Wallis H; U: Mann-Whitney U; t: t-testz = standardized test statistic.
* Distributions of ABILHAND measures scores were not similar between groups (assessed by visual inspection of a boxplot). As such, statistical analysis was carried by evaluation of mean ranks. Medians and IQR are reported for easier interpretation.
# Pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. This post hoc analysis revealed statistically significant differences in the mean ranks of ABILHAND scores between diffuse SSc and VEDOSS (p = 0.034), but not between any other two groups.
+Age categories were calculated by identifying quartiles and rounding to the nearest multiple of 5.

TABLE V. Correlations between ABILHAND-SSc score and other patient-reported outcomes

		Coefficient	p-value
EQ-5D-5L	Index	ρ=0.684	p<0.0001
	EQ-VAS	r=0.463	p=0.001
SF-36v2	Physical functioning	ρ=0.582	p<0.0001
	Physical role functioning	ρ=0.433	p=0.001
	General health perceptions	r=0,569	p<0.0001
	Vitality	r=0.489	p<0.0001
	Social role functioning	ρ=0.247	p=0.075
	Emotional role functioning	ρ=0.426	p=0.002
	Mental health	r=0.378	p=0.005
SHAQ	HAQ-DI	ρ=-0.699	p<0.0001
	SHAQ Raynaud's VAS	ρ=-0.361	p=0.009
	SHAQ Finger ulcer VAS	ρ=-0.247	p=0.078
	SHAQ Overall disease everity VAS	ρ=-0.541	p<0.0001
ScleroID	Total score	ρ=-0.698	P<0.0001

r – Pearson's coefficient; VAS – visual analogue scale; ρ – spearman's coefficient

differential item functioning (DIF), particularly by age and subtype, suggesting that certain tasks may be perceived differently depending on patient characteristics. Although these differences did not compromise the overall measurement model, they highlight the importance of evaluating DIF in PROMs used across heterogeneous clinical phenotypes.

Test–retest reliability was evaluated in a subset of 13 patients (approximately one quarter of the total sample) who were considered clinically stable. This limited number reflects logistical challenges and suboptimal adherence to the retesting procedure, despite efforts to minimize patient burden. Also of note, the test-retest reliability in VEDOSS or SSc sine scleroderma patients is unclear, as none of these patients participated in the retest. Nevertheless, the observed ICC supports good temporal stability of the scale. The modest sample size and potential variability in perceived hand function over short intervals may have contributed to some measurement variability.

This study has important strengths, including the use of a Rasch-based approach, involvement of patients and SSc experts in translation and adaptation, and the inclusion of multiple PROMs for external validation. Nevertheless, several limitations must be acknowledged. First, the study was conducted in a single tertiary care centre, which may limit generalizability to other Portuguese-speaking regions or care settings. Second, although educational diversity was observed, socioeconomic status was not systematically recorded, precluding analysis of its impact on item comprehension or response. Third, the small number of VEDOSS and sine SSc patients limits the interpretation of subgroup comparisons and the assessment of their impact on test performance. Finally, the cross-sectional design precluded analysis of responsiveness to clinical change over time.

CONCLUSION

The European Portuguese version of the ABILHAND-SSc demonstrated good psychometric properties, including feasibility, reliability, and construct validity. It is a valuable tool for assessing hand function in patients with systemic sclerosis and may support clinical monitoring, patient-centred care, and future research on functional outcomes in this population. Future work should validate its responsiveness in longitudinal studies and confirm its applicability in broader, multicentre cohorts.

REFERENCES

1. Varga J, Abraham D. Systemic sclerosis: a prototypic multisystem fibrotic disorder. J Clin Invest. 2007 Mar;117(3):557-67. https://doi.org/10.1172/JCI31139
2. Almeida C, Almeida I, Vasconcelos C. Quality of life in systemic sclerosis. Autoimmun Rev. 2015 Dec;14(12):1087-96. https://doi.org/10.1016/j.autrev.2015.07.012

3. Pauling JD, Caetano J, Campochiaro C, et al. Patient-reported outcome instruments in clinical trials of systemic sclerosis. J Scleroderma Relat Disord. 2020 Jun;5(2):90-102. https://doi.org/10.1177/2397198319886496
4. Ahrens HC, Siegert E, Tomsitz D, et al. Digital ulcers score: a scoring system to assess digital ulcers in patients suffering from systemic sclerosis. Clin Exp Rheumatol. 2016 Sep-Oct;34 Suppl 100(5):142-147. Epub 2016 Oct 13. PMID: 27749240.
5. Sandler RD, Matucci-Cerinic M, Hughes M. Musculoskeletal hand involvement in systemic sclerosis. Semin Arthritis Rheum. 2020 Apr;50(2):329-334. https://doi.org/10.1016/j.semarthrit.2019.11.003
6. Penta M, Thonnard JL, Tesio L. ABILHAND: a Rasch-built measure of manual ability. Arch Phys Med Rehabil. 1998 Sep;79(9):1038-42. https://doi.org/10.1016/S0003-9993(98)90167-8
7. Vanthuyne M, Smith V, Arat S, et al. Validation of a manual ability questionnaire in patients with systemic sclerosis. Arthritis Rheum. 2009 May 15;61(5):695-703. doi: 10.1002/art.24426. PMID: 19405012. https://doi.org/10.1002/art.24426
8. Patrício R. Translation and cross-cultural adaptation of the manual ability questionnaire ABILHAND in systemic sclerosis into european portuguese language. University of Coimbra; 2022. [cited 2025 Jan 3]. Available from: https://hdl.handle.net/10316/102341
9. van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. Ann Rheum Dis. 2013;72:1747-1755. https://doi.org/10.1136/annrheumdis-2013-204424
10. Avouac J, Fransen J, Walker U, et al. Preliminary criteria for the very early diagnosis of systemic sclerosis: results of a Delphi Consensus Study from EULAR Scleroderma Trials and Research Group. Ann Rheum Dis. 2011;70:476-481. https://doi.org/10.1136/ard.2010.136929
11. Ferreira PL. Development of the Portuguese version of MOS SF-36. Part II --Validation tests. Acta Med Port. 2000;13:119-127.X
12. Ferreira PL, Antunes P, Ferreira LN, Pereira LN, Ramos-Goñi JM. A hybrid modelling approach for eliciting health state preferences: the Portuguese EQ-5D-5L value set. Quality of Life Research. 2019;28:3163-3175. https://doi.org/10.1007/s11136-019-02226-5
13. Steen VD, Medsger TA. The value of the health assessment questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. Arthritis Rheum. 1997;40:1984-1991. https://doi.org/10.1002/art.1780401110
14. Genrinho I, Ferreira PL, Santiago T, et al. Validation of the Portuguese Version of the Scleroderma Health Assessment Questionnaire. Int J Environ Res Public Health. 2023;20:7062. https://doi.org/10.3390/ijerph20227062
15. Becker MO, Dobrota R, Garaiman A, et al. Development and validation of a patient-reported outcome measure for systemic sclerosis: the EULAR Systemic Sclerosis Impact of Disease (SclerоID) questionnaire. Ann Rheum Dis. 2022;81:507-515. https://doi.org/10.1136/annrheumdis-2021-220702
16. Vieira M. Translation and cross-cultural adaptation of the EULAR systemic sclerosis impact of disease (ScleroID) into European Portuguese language. University of Coimbra; 2022. [cited 2025 Jan 3]. Available from: https://hdl.handle.net/10316/102447

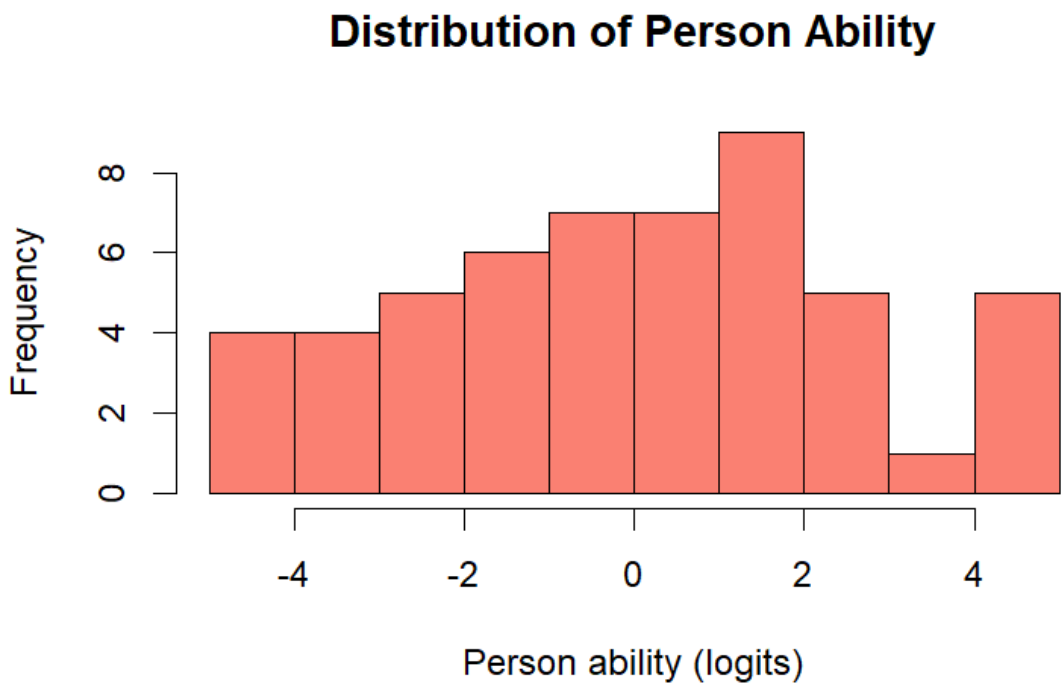
SUPPLEMENTARY MATERIAL

APPENDIX I. European Portuguese Version of the SSc-adapted ABILHAND questionnaire

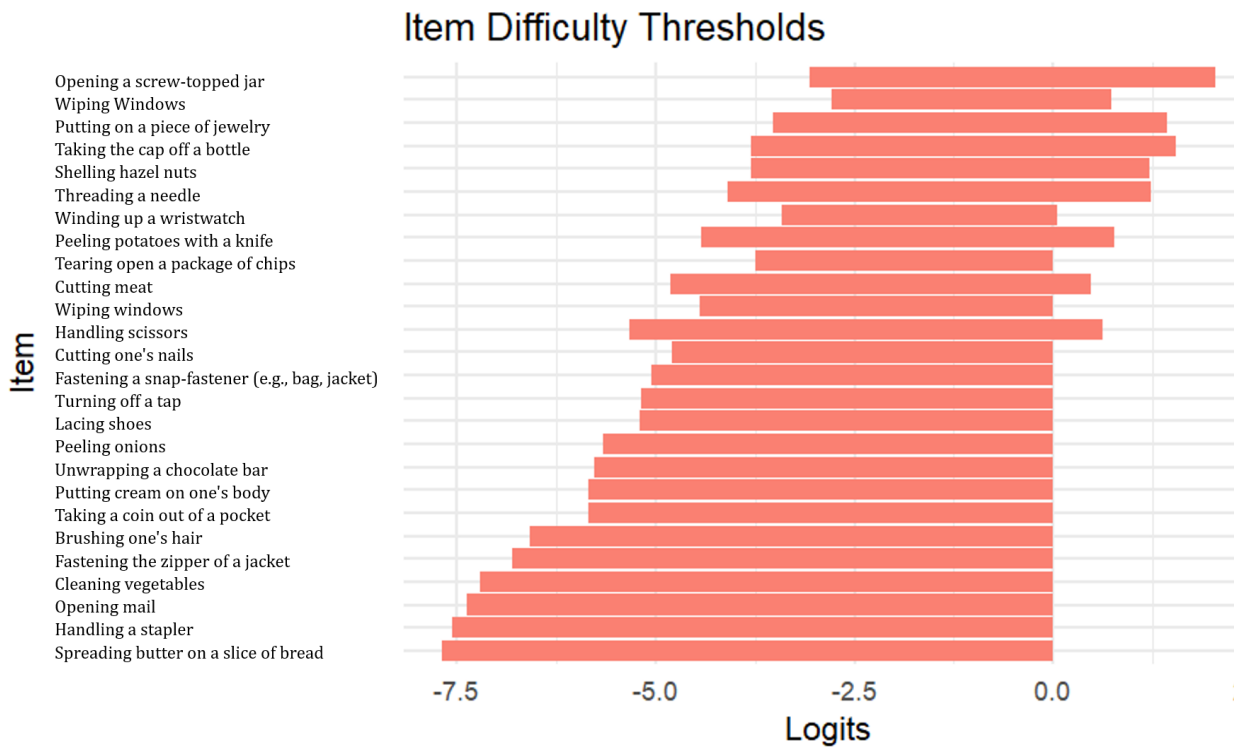
ABILHAND – Avaliação da habilidade manual

Qual o seu grau de dificuldade nas seguintes atividades?	Impossível	Difícil	Fácil	OBS
1. Enfiar uma linha na agulha				
2. Colocar joias em si próprio/a				
3. Cortar carne				
4. Usar uma tesoura				
5. Tirar uma tampa de uma garrafa				
6. Tirar uma moeda do bolso				
7. Cortar as unhas				
8. Desembrulhar um chocolate				
9. Limpar as janelas				
10. Fechar uma torneira				
11. Apertar os atacadores dos sapatos				
12. Usar um agrafador				
13. Abrir um envelope				
14. Descascar cebolas				
15. Dar corda a um relógio				
16. Descascar batatas com uma faca				
17. Desenroscar a tampa dum frasco				
18. Fechar o fecho de um casaco				
19. Lavar vegetais				
20. Barrar manteiga numa fatia de pão				
21. Espalhar creme no corpo				
22. Abrir um pacote de batatas fritas com as mãos				
23. Pentear o cabelo				
24. Fechar um botão de mola (por exemplo, de um saco, blusão)				
25. Descascar castanhas				
26. Abotoar as calças				

ABILHAND measurement: Analysis of the answers is via a Rasch model of online analysis, which converts the raw scores into a linear measure. It can be found at: https://www.rehab-scales.org/



Appendix II. Distribution of patient abilities estimated by the Rasch model
Histogram showing the distribution of patient ability estimates in logits. The distribution confirms that most patients had manual ability levels above the average item difficulty.



Appendix III. Item difficulty thresholds (logits)
Bar plot of item thresholds (Category 1 and 2) for each ABILHAND-SSc item on the logit scale. This provides a single difficulty estimate per item, reflecting its overall location on the linear logit scale. More negative values indicate easier items (tasks accessible to patients with lower ability), while more positive values indicate harder items (tasks requiring greater manual ability).

APPENDIX IV. Item fit statistics and thresholds of the ABILHAND-SSc (Rasch model)

Item	Threshold	Infit	Infit_t	Infit_p	Misfitting
Threading a needle	Threshold 1	1.828	2.296	0.0217	No
	Threshold 2	1.34	1.623	0.1045	No
Putting on a piece of jewelry	Threshold 1	1.267	1.002	0.3163	No
	Threshold 2	0.942	-0.233	0.8159	No
Cutting meat	Threshold 1	1.174	0.526	0.5994	No
	Threshold 2	1.118	0.649	0.5169	No
Handling scissors	Threshold 1	1.28	0.648	0.517	No
	Threshold 2	0.706	-1.683	0.0914	No
Taking the cap off a bottle	Threshold 1	0.63	-1.438	0.1505	No
	Threshold 2	0.963	-0.135	0.8919	No
Taking a coin out of a pocket	Threshold 1	0.896	-0.041	0.9673	No
	Threshold 2	0.691	-1.582	0.1373	No
Cutting one's nails	Threshold 1	1.134	0.482	0.6304	No
	Threshold 2	2.02	0.155	0.8768	No
Unwrapping a chocolate bar	Threshold 1	1.092	0.345	0.7299	No
	Threshold 2	0.639	-1.763	0.0778	No
Wiping windows	Threshold 1	0.862	-0.49	0.6242	No
	Threshold 2	1.224	1.141	0.2539	No
Turning off a tap	Threshold 1	0.762	-0.808	0.4189	No
	Threshold 2	1.618	2.267	0.0234	No
Lacing shoes	Threshold 1	0.596	-1.524	0.1274	No
	Threshold 2	0.703	-1.525	0.1274	No
Handling a stapler	Threshold 1	1.086	0.371	0.7108	No
	Threshold 2	0.541	-2.275	0.0229	No
Opening mail	Threshold 1	0.921	0.011	0.9909	No
	Threshold 2	0.666	-1.4	0.1605	No
Peeling onions	Threshold 1	0.973	0.108	0.9144	No
	Threshold 2	0.947	-0.194	0.8466	No
Winding up a wristwatch	Threshold 1	1.216	0.834	0.4041	No
	Threshold 2	1.087	0.434	0.664	No
Peeling potatoes with a knife	Threshold 1	1.102	0.382	0.7022	No
	Threshold 2	1.076	0.443	0.6584	No
Opening a screw-topped jar	Threshold 1	0.799	-0.764	0.4449	No
	Threshold 2	0.837	-0.703	0.4826	No
Fastening the zipper of a jacket	Threshold 1	1.077	0.362	0.7172	No
	Threshold 2	0.723	-1.345	0.1787	No
Cleaning vegetables	Threshold 1	1.06	0.271	0.7857	No
	Threshold 2	1.171	0.713	0.4759	No
Spreading butter on a slice of bread	Threshold 1	1.088	0.322	0.7469	No
	Threshold 2	1.169	0.668	0.5041	No
Putting cream on one's body	Threshold 1	0.924	0.011	0.9909	No
	Threshold 2	0.939	-0.218	0.8275	No
Tearing open a package of chips	Threshold 1	1.206	0.8	0.4235	No
	Threshold 2	1.129	0.653	0.5138	No
Brushing one's hair	Threshold 1	1.303	0.693	0.4878	No
	Threshold 2	1.024	0.172	0.8636	No
Fastening a snap-fastener (e.g., bag, jacket)	Threshold 1	0.896	-0.148	0.8199	No
	Threshold 2	0.796	-1.022	0.3068	No
Shelling hazel nuts	Threshold 1	0.975	-0.018	0.9858	No
	Threshold 2	0.986	-0.025	0.9804	No
Buttoning up trousers	Threshold 1	0.849	-0.263	0.7922	No
	Threshold 2	0.836	-0.774	0.4391	No

Each item includes two thresholds: Threshold 1 corresponds to the transition from response category 0 ("Impossible") to 1 ("Difficult"), and Threshold 2 to the transition from category 1 to 2 ("Easy"). Infit mean square values, standardized t-statistics (Infit_t), and p-values are reported for each threshold. No threshold showed misfit based on t-statistics ($|t| > 2.5$) or Bonferroni-adjusted p-values (< 0.002).

Appendix V. Internal consistency statistics for the ABILHAND-SSc items

Item	Sex (p)	Age group (p)	Disease duration (p)	SSc subtype (p)	Skin thickening (prox. to MCPs) (p)
Threading a needle	0.9792	0.1558	0.2300	0.3929	**< 0.0001**
Putting on a piece of jewelry	0.5158	0.6338	0.3845	0.0580	0.1935
Cutting meat	0.1118	0.370	0.4962	0.0049	0.0055
Handling scissors	0.0335	0.2329	0.3647	**< 0.0001**	0.0901
Taking the cap off a bottle	0.2102	0.0109	0.5101	0.0347	0.4803
Taking a coin out of a pocket	0.8659	0.2411	0.8867	0.1571	0.0077
Cutting one's nails	0.1248	0.0361	0.0907	0.0066	0.1484
Unwrapping a chocolate bar	0.0310	0.0837	0.5136	0.1769	0.6721
Wiping windows	0.2864	0.0430	0.0764	0.9251	**0.0008**
Turning off a tap	0.1812	0.2398	0.1241	0.0071	0.0227
Lacing shoes	0.6966	**0.0011**	0.0073	**0.0015**	0.0780
Handling a stapler	0.1869	**0.0002**	0.3104	0.4854	0.0540
Opening mail	0.3394	**0.0002**	0.3019	0.0059	0.0162
Peeling onions	0.6054	0.2530	0.8455	**0.0016**	0.0936
Winding up a wristwatch	0.2550	0.0247	0.1671	**0.0007**	0.0166
Peeling potatoes with a knife	0.0596	0.9487	0.5612	0.0849	0.2536
Opening a screw-topped jar	0.4818	0.0045	0.2030	0.0074	0.0033
Fastening the zipper of a jacket	0.6243	0.1318	0.9279	0.1899	0.1243
Cleaning vegetables	0.1618	0.0897	0.1803	**0.0014**	0.1450
Spreading butter on a slice of bread	0.9730	0.1754	0.3799	0.1205	**0.0012**
Putting cream on one's body	0.1842	0.0357	0.0566	0.3778	0.0215
Tearing open a package of chips	0.3038	0.0749	0.4731	0.0125	**< 0.0001**
Brushing one's hair	0.8137	0.1046	0.2004	**0.0008**	0.6441
Fastening a snap-fastener (e.g., bag, jacket)	0.8801	0.0589	0.5249	0.0049	0.0150
Shelling hazel nuts	0.3440	0.0032	0.1597	0.0331	0.3222

This table displays the p-values from the two-way ANOVA assessing Differential Item Functioning (DIF) across four covariates: sex, age group, disease duration, and SSc subtype. Person ability was stratified into tertiles and included as a factor. Bonferroni correction was applied ($p < 0.002$). Lower p-values indicate greater evidence of DIF.

APPENDIX VI. Internal consistency statistics for the ABILHAND-SSc items

Item	Corrected Item–Total Correlation	Cronbach's Alpha if Item Deleted
Threading a needle	0.462	0.967
Putting on a piece of jewelry	0.662	0.966
Cutting meat	0.654	0.966
Handling scissors	0.727	0.965
Taking the cap off a bottle	0.758	0.965
Taking a coin out of a pocket	0.821	0.964
Cutting one's nails	0.756	0.965
Unwrapping a chocolate bar	0.781	0.965
Wiping windows	0.731	0.965
Turning off a tap	0.632	0.966
Lacing shoes	0.850	0.964
Handling a stapler	0.827	0.964
Opening mail	0.758	0.965
Peeling onions	0.717	0.965
Winding up a wristwatch	0.702	0.965
Peeling potatoes with a knife	0.681	0.965
Opening a screw-topped jar	0.770	0.965
Fastening the zipper of a jacket	0.791	0.965
Cleaning vegetables	0.583	0.966
Spreading butter on a slice of bread	0.569	0.966
Putting cream on one's body	0.739	0.965
Tearing open a package of chips	0.679	0.966
Brushing one's hair	0.661	0.966
Fastening a snap-fastener (e.g., bag, jacket)	0.796	0.964
Shelling hazel nuts	0.717	0.965
Buttoning up trousers	0.792	0.964

For each of the 26 items in the ABILHAND-SSc, the corrected item–total correlation and the Cronbach's alpha if the item was deleted are reported. Corrected item–total correlations > 0.3 and minimal changes in Cronbach's alpha upon item deletion support item homogeneity and overall internal consistency of the scale.