

The HAND-Q: Psychometrics of a New Patient-reported Outcome Measure for Clinical and Research Applications

Kyra L. Sierakowski, MD, PhD*
 Nicola R. Dean, MBChB, PhD*
 Kathleen Evans Sanchez, MD*
 Philip A. Griffin, MBBS, FRACS*
 Justin C.R. Wormald, MRes†
 Jeremy N. Rodrigues, PhD,
 FRCS(Plast)†
 Olivier Mares, MD‡
 Jussi P. Repo, MD, PhD§
 Sina M. Hulkkonen, MD, PhD¶
 Neil V. Shah, MD, MS||
 Steven Koehler, MD, FAAOS||
 Gregory I. Bain, MBBS, PhD**
 Stefan J. Cano, PhD††
 Andrea L. Pusic, MD, MHS‡‡
 Don Lalonde, MD§§
 Anne F. Klassen, DPhil¶¶

Background: The perspective of the patient in measuring the outcome of their hand treatment is of key importance. We developed a hand-specific patient-reported outcome measure to provide a means to measure outcomes and experiences of care from the patient perspective, that is, HAND-Q.

Methods: Data were collected from people with a broad range of hand conditions in hand clinics in six countries between April 2018 and January 2021. Rasch measurement theory analysis was used to perform item reduction and to examine reliability and validity of each HAND-Q scale.

Results: A sample of 1277 patients was recruited. Participants ranged in age from 16 to 89 years, 54% were women, and a broad range of congenital and acquired hand conditions were represented. Rasch measurement theory analysis led to the refinement of 14 independently functioning scales that measure hand appearance, health-related quality of life, experience of care, and treatment outcome. Each scale evidenced reliability and validity. Examination of differential item functioning by age, gender, language, and type of hand condition (ie, nontraumatic versus traumatic) confirmed that a common scoring algorithm for each scale could be implemented.

Conclusions: The HAND-Q was developed following robust psychometric methods to provide a comprehensive modular independently functioning set of scales. HAND-Q scales can be used to assess and compare evidence-based outcomes in patients with any type of hand condition. (*Plast Reconstr Surg Glob Open* 2022;10:e3998; doi: [10.1097/GOX.0000000000003998](https://doi.org/10.1097/GOX.0000000000003998); Published online 31 January 2022.)

INTRODUCTION

Strength, sensibility, and range of motion are commonly used as metrics for gauging the success of hand

surgery or other treatments. Less common to routine clinical practice is the incorporation of the patient perspective, which is measurable using patient-reported outcome measures (PROMs). There are currently a range of hand-specific PROMs that can be divided into regional anatomical PROMs designed for the upper limb or hand, and PROMs

*From the *Department of Plastic and Reconstructive Surgery, College of Medicine and Public Health, Flinders University, Adelaide, South Australia, Australia; †Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Nuffield Orthopaedic Centre, Windmill Road, Oxford, United Kingdom; ‡Department of Traumatology and Orthopedic Surgery, Centre Hospitalo-Universitaire (CHU) Carêmeau, Nîmes, France; §Department of Orthopaedic and Trauma Surgery, Unit of Musculoskeletal Disease, Tampere University Hospital, Tampere, Finland; ¶Department of Hand Surgery, Helsinki University Hospital and University of Helsinki, Finland; ||Department of Orthopaedic Surgery and Rehabilitation, State University of New York (SUNY) Downstate Medical Center, Brooklyn, N.Y.; **Department of Orthopaedic Surgery, Flinders University, Adelaide, South Australia, Australia; ††Modus Outcomes, Letchworth Garden City, United Kingdom; ‡‡Department of Surgery, Patient-Reported Outcome, Value and Experience (PROVE) Center, Harvard Medical School, Brigham and Women's Hospital, Boston, Mass.; §§Division of Plastic Surgery, Dalhousie University, Saint John, New Brunswick, Canada; and ¶¶McMaster University, Hamilton, Ontario, Canada. The HAND-Q study has been funded by research grants from the Australasian Foundation for Plastic Surgery—Foundation Plastic Surgery and Reconstructive Surgical (PRS) Research Grant, and the Royal Australasian College Surgeons Small Project Grant. No other sources of funds supported this work. No products, devices or drugs were used in this study.*

Received for publication August 23, 2021; accepted October 27, 2021.

Copyright © 2022 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 \(CCBY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal. DOI: [10.1097/GOX.0000000000003998](https://doi.org/10.1097/GOX.0000000000003998)

Disclosure: Drs. Sierakowski, Dean, Pusic, and Klassen are co-developers of the HAND-Q and will receive a share of any license revenue on the inventor sharing policies of the institutions where the HAND-Q was developed. Dr Klassen is owner of EVENTUM Research which provides consulting services to the pharmaceutical industry. The other authors have no financial interest to declare.

Related Digital Media are available in the full-text version of the article on www.PRSGlobalOpen.com.

designed for specific hand conditions. The development and psychometric properties of current available instruments for hand conditions are the subject of a growing number of systematic reviews.¹⁻⁴ The findings show that few current PROMs fulfill international development and validation guidelines.¹⁻⁴ For example, the Disabilities of the Arm, Shoulder and Hand (DASH)⁵ was developed without the input of patients with hand conditions, which is a necessity in the development of instruments measuring the perspective of the patient.⁴ Furthermore, most PROMs have been developed using psychometric methods that are now outdated, resulting in ordinal measurement that results in unequal distances between scores of the scale, like a ruler without consistency between measurement points. As a result, these PROMs are not ideal for measuring change over time in research studies or with individual patients, since the application of scores over the full breadth of the scale is not equal.^{4,6,7} PROMs developed using Rasch measurement theory (RMT) analysis,⁸ or item response theory,⁹ produce interval data with equal distances between scores on a scale. Such instruments are mathematically sound and provide a means to accurately measure change over time in research studies and in patient care.

Of the PROMs commonly used in hand research, the DASH,⁵ QuickDASH (qDASH),¹⁰ and Michigan Hand Outcomes Questionnaire¹¹ stand out as having the most published psychometric properties. But even these PROMs have incomplete evidence to support their use in hand surgery research and clinical practice based on contemporary PROM standards, such as put forward by the COnsensus-based Standards for the selection of health Measurement INSTRUMENTS (COSMIN).^{12,13}

Hand conditions are extremely common and have a major impact on patients' lives. Since patients are the best source of information on how they function and feel, it is vital to include their perspective in the assessment of outcomes. A comprehensive PROM for hand conditions developed using a modern psychometric approach is required. Such a PROM needs to address the varied impacts of hand conditions on an individual, with scales designed to measure unidimensional concepts in order to allow for ease of interpretation of scores.

To address the need for a comprehensive hand-specific PROM, our team recently developed the HAND-Q. We interviewed 62 people with a range of conditions, including carpal tunnel syndrome, Dupuytren disease, trigger finger, osteoarthritis, rheumatoid arthritis, and injuries of various kinds. Analysis of the qualitative data led to the development of a conceptual framework covering important outcomes and care experiences. Key concepts were included in scales that were refined through cognitive debriefing interviews with 20 patients, and a survey of 25 healthcare professions with expertise in treating hand conditions. Findings from phase 1 are published elsewhere.¹⁴ Following guidelines,¹⁵ the scales were translated and culturally adapted into French and Finnish for inclusion in a phase 2 multinational field-test study.

The aim of this paper was to describe the psychometric findings from the phase 2 field-test study that used a modern psychometric method (ie, RMT analysis⁸) to

Takeaways

Question: Will the HAND-Q prove to be reliable and valid in an international field-test study?

Findings: In total, 1277 people with a broad range of hand conditions from six countries completed the HAND-Q. A modern psychometric analysis provided evidence to support the use of 14 scales that measure hand appearance, health-related quality of life, experience of care, and treatment.

Meaning: This study provides researchers and clinicians with a new questionnaire measuring outcomes that matter to people with hand conditions.

examine each HAND-Q scale, and to remove items and scales that were redundant or exhibited poor psychometric performance. We conducted additional tests of construct validity, including correlations between scores on HAND-Q scales, and examination of relationships between HAND-Q scores and clinical variables, that is, visibility and severity of the hand condition, and need for surgery.

METHODS

Research Ethics

Research ethics board approval was obtained for the phase 2 field-test study from seven sites in six countries, namely Canada, Australia, the United Kingdom, the United States, France and Finland (see figure, **Supplemental Digital Content 1**, which shows the name of the ethics board to approve the study in each collaborating site, <http://links.lww.com/PRSGO/B894>). The first board to grant approval was the Southern Adelaide Clinical Human Research Ethics Committee, Adelaide.

Sample and Recruitment

The phase 2 field-test study took place between April 2018 and January 2021. Participants were aged 16 years and older with a hand condition for which they attended an outpatient clinic. Recruitment took place in clinics using either paper-based forms or electronic data capture via a tablet device. All data were entered into a REDCap survey¹⁶ hosted at Flinders University, Adelaide, South Australia, Australia.

Survey

The REDCap survey asked demographic (age, gender, education, and occupation) and clinical questions, including the side of the hand condition, hand dominance, type of condition and its severity (mild, moderate, and severe), how long the participant has had the condition, and whether surgery was needed (yes, no, and not sure). **Table 1** shows scale characteristics for the 20 HAND-Q scales that were field-tested, and the branching logic used in the REDCap survey to ensure that each scale was completed by participants for whom it was relevant.

Table 1. HAND-Q Scales Tested in Field-test Study, including Branching Logic

Domain	Scale	Response Options	Recall Period	Items	Branching Logic	
Appearance HRQL	Appearance	Satisfaction	Now	30	None	
	Acceptance	Agreement	None	7	Hand problem lasting ≥ 6 mo	
	Function	Difficulty	Past week	35	None	
	Life impact	Severity	Past week	11	None	
	Psychological	Frequency	Past week	19	None	
	Sexual	Bothered	None	9	Hand problem affects sex life	
	Sleep	Frequency	Past week	8	Interfered with ≥ 1 night in past week	
	Social	Agreement	Past week	13	None	
	Symptoms	Severity	Past week	22	None	
	Work	Agreement	None	11	Worked in job in past 3 mo	
	Experience	Anesthesia	Bothered	None	14	Had surgery ≤ 7 mo ago
		Anesthesia symptoms (post)	Severity	None	13	Had surgery ≤ 7 mo ago
		Awake procedure	Satisfaction	None	17	Had surgery ≤ 7 mo ago and had local anesthesia
Hand clinic		Agreement	Recent appointments	13	Saw ≥ 1 in past 3 mo and not first appointment	
Hand therapist		Agreement	Recent appointments	19	Saw ≥ 1 in past 3 mo and not first appointment	
Information		Satisfaction	None	20	Had surgery ≤ 7 mo ago	
Office staff		Agreement	Recent appointments	14	Saw ≥ 1 in past 3 mo and not first appointment	
Treatment	Surgeon	Agreement	Recent visit	25	Saw ≥ 1 in past 3 mo and not first appointment	
	Outcome	Agreement	Recent treatment	9	Saw ≥ 1 in past 3 mo and not first appointment	
	Splint	Satisfaction	Recent splint	12	Saw ≥ 1 and used splint/brace in past 3 mo	

Analysis

Data for each language version were merged in SPSS Version 26.0 (IBM Corporation, Armonk, N.Y., for Windows/Apple Mac). RMT analysis was performed using RUMM2030 software (RUMM version 2030, RUMM Laboratory Pty Ltd, Duncraig, Western Australia, 1998–2021) and the unrestricted Rasch model for polytomous data. Scales that comply to this model have a series of items that measure an amount of a single concept.^{8,17} A range of statistical and graphical tests were used to determine how well the data fit the Rasch model; specifically, the following tests were performed:

Thresholds for Item Response Options

Item response options were investigated to check whether item thresholds were ordered appropriately.¹⁸

Item Fit Statistics

Various fit indicators (item characteristic curve), log residuals (item–person interaction), and chi-square (item–trait interaction) were considered in conjunction with each other. Fit residual indicates if the item under or overdiscriminates compared with other items, and provide information about redundancy. Ideal fit residuals are between -2.5 and $+2.5$, with a nonsignificant chi-squared value after Bonferroni adjustment.¹⁸ The sample was adjusted to 500 for tests of item fit.¹⁷

Targeting

Ideally, items of a scale should reflect all levels of the measured concept experienced by the sample.¹⁷ Both graphical (person-item threshold distribution) and statistical (proportion of sample not captured by the range of the scale) tests were examined to establish if the items were adequately spread for each scale.

Differential Item Function

Sample characteristics were examined for differential item function (DIF), including age group (16–39, 40–59, and ≥ 60 years), gender (men versus women),

type of hand condition [nontraumatic (ie, elective) versus traumatic] and language (English versus other). DIF testing examines whether there is possible response bias between subgroups. Scales with a minimum data set of 150 responses were examined for DIF. A random sample of equal sized groups was used for the analysis, which was repeated three times. If DIF was found for any item in any of the three random sample analyses, the items with DIF were split on the sample characteristic. A Pearson correlation was then performed on the original and the new person locations to examine any impact of DIF on the scoring of the scale.¹⁸

Reliability

Person separation index (PSI) and Cronbach alpha values were used to examine reliability of each scale.²⁰ The values range between 0 and 1, with a higher value indicating higher reliability. Coefficients of 0.70 or greater were considered adequate.²¹

Local Independence

Items were examined for residual correlations over 0.30, which were taken to indicate that answers for an item may depend on answers to another item.^{22–24} Subtesting was undertaken to determine any change in the PSI for scales with residual correlations above 0.30.¹⁹

Unidimensionality

The final set of items for each scale underwent principal component analysis in SPSS, version 26.0. It was hypothesized that items of each scale would relate to a single factor with loadings greater than 0.70.²⁵

Construct Validity

Rasch logits were used to transform participant scores for each scale from 0 (worst) to 100 (best). Normality was assessed using Kurtosis and Skewness, and nonparametric statistics were used if distributions were nonnormal (ie, outside of -2 to 2).²⁵ The scores were used to test four hypotheses to examine construct validity as follows:

1. HAND-Q scale scores within domains [eg, health-related quality of life (HRQL)] would correlate more strongly with each other than with scale scores from other domains.
2. HAND-Q scores for the scales measuring hand appearance would be lower for participants with a visible (ie, Dupuytren contracture, rheumatoid arthritis, and oosteoarthritis) versus an invisible (ie, carpal tunnel syndrome) hand condition.
3. HAND-Q scores for scales in the HRQL and treatment outcome domains would be incrementally lower for increasing severity (ie, mild, moderate, and severe) of the hand condition.
4. HAND-Q scores would be lower on the HRQL and treatment outcome scales for participants who needed versus did not need further surgery.

RESULTS

Demographics

Characteristics of the 1277 participants are shown in [Table 2](#). Recruitment was highest in Australia and the United States, with smaller contributions from the United Kingdom, Canada, Finland, and France. The mean age of the sample was 50 (SD = 17) and ranged from 16 to 89 years. Most data were collected using the English version of the HAND-Q with other contributions in Finnish and French. The most common hand conditions were carpal tunnel syndrome, soft-tissue injury, and fractures. Nontraumatic conditions were more common than traumatic conditions. Most participants (N = 812, 63.6%) had a condition that affected their dominant hand. A total of 776 (61%) participants worked at a job in the past 3 months, and most of those who worked (N = 548, 71%) did so full time.

RMT Analysis

In the RMT analysis, data collected for seven scales (ie, acceptance, sleep, social, work, anesthesia, anesthesia symptoms, and awake procedure) were found to be incompatible with RMT analysis and were dropped. These scales had multiple items with disordered thresholds. After we rescored each scale's items to reduce by one threshold, and dropped items with poor fit, scale reliability was low (PSI values < 0.70).

RMT analysis provided evidence of reliability and validity for the remaining 14 scales, that is, seven HRQL scales, five experience of care scales, and two treatment outcome scales. The HRQL scales include two for hand appearance, that is, one measures hand appearance in general, and the other measures age-related hand appearance. The total number of items in the 14 scales was reduced from 238 to 133. The clinic scale had four items with disordered thresholds. After reducing response options by rescoring across the “definitely disagree” and “somewhat disagree” categories, item thresholds were ordered. The RMT analysis used the rescored data.

Figure, Supplemental Digital Content 2, shows the item fit and DIF statistics (<http://links.lww.com/PRSGO/>

Table 2. Characteristics of the 1277 Field-test Study Participants

		N	%	
Country	Australia	446	35	
	Canada	85	7	
	Finland	184	14	
	France	82	6	
	UK	98	8	
	USA	382	30	
Language	English	1011	79	
	Other	266	21	
Gender	Male	574	45	
	Female	680	53	
	Other	6	1	
	Missing	17	1	
Age, y	16–39	360	28	
	40–59	507	40	
	60 +	392	31	
	Missing	18	1	
Education	Primary school	65	5	
	High school	564	44	
	Further education (college, university, or similar)	572	45	
	Other	54	4	
	Missing	22	2	
	Hand condition	Carpal tunnel syndrome	183	14
		Dupuytren's disease	36	3
Trigger finger		70	5	
Osteoarth		62	5	
Rheumatoid arthritis		13	1	
Other nerve compression		23	2	
Ganglion		18	1	
Trauma (injury and fracture)		519	38	
Multiple nontraumatic conditions		117	9	
Mixed nontraumatic and traumatic conditions		65	5	
Other		171	13	
Type of condition		Nontraumatic	687	54
		Traumatic	525	41
	Both	65	5	
Need further surgery	Yes	95	7	
	No	1170	92	
	Missing	12	1	
Severity of condition	Mild	202	16	
	Moderate	521	41	
	Severe	521	41	
	Missing	33	3	

B895). All 133 items had nonsignificant X^2 P values after Bonferroni adjustment. Item fit was within ± 2.5 for 102 items. A total of 77 items from eight scales were tested for DIF. Of these, 40 items demonstrated DIF, including 12 when tested by gender, 15 when tested by age-group, 19 when tested by type of hand condition, and 21 when tested by language. Pearson correlations between person locations before and after splitting the items for DIF showed minimal impact on scoring (all correlations > 0.99).

Scale-level findings are summarized in [Table 3](#). Data fit the Rasch model for 11 scales, which had nonsignificant P values, with marginal misfit for the sexual, clinic, and splint scales. The proportion of participants to score within the range provided by each of the HRQL scales ranged from 94% (symptoms) to 76% (sexual). Targeting is further illustrated in [Figure 1](#), which shows the findings for the two physical scales (ie, function and symptoms) as examples.

The figures show the distribution of person measurement (top histograms) and item locations (lower histograms), with higher (better) scores to the right. The person measurements are shown separately by type of hand condition

Table 3. Rasch Measurement Theory and Scale-level Statistics

Domain	Scale	N Final Items	N Completed Scale	N in RMT	% Scored on Scale	χ^2	DF	<i>P</i>	PSI +ext	PSI -ext	CA +ext	CA -ext	% Floor	% Ceiling	% Missing Data
Appearance	Appearance	10	1190	923	78	105.54	90	0.13	0.86	0.87	0.95	0.91	0.8	19.3	12.4
	Appearance: age-related	10	1188	952	80	88.50	90	0.52	0.86	0.87	0.94	0.91	0.8	16.7	11.4
HRQL	Function	15	1213	1031	85	161.34	135	0.06	0.92	0.93	0.97	0.96	4.3	9.2	12.9
	Life impact	8	1174	1031	88	84.14	72	0.15	0.83	0.81	0.91	0.87	2.1	9.6	5.6
	Psychological	10	1192	1031	87	92.79	90	0.40	0.82	0.83	0.92	0.90	0.4	12.8	4.5
	Sexual	7	445	313	76	42.69	28	0.04	0.87	0.86	0.95	0.90	4.7	22.9	6.5
	Symptoms	10	1183	1111	94	55.79	80	0.98	0.84	0.83	0.90	0.87	1.4	4.3	8.9
Experience	Clinic	10	500	258	52	50.57	30	0.01	0.67	0.77	0.93	0.84	0.4	44.4	10.0
	Doctor	10	634	181	29	31.44	20	0.05	0.51	0.81	0.96	0.91	0.6	67.0	5.7
	Hand therapist	10	249	65	26	23.38	20	0.27	0.71	0.90	0.98	0.94	0.8	70.3	6.4
	Information	10	430	214	50	28.9	20	0.09	0.75	0.87	0.96	0.93	0.7	45.1	7.2
Treatment	Office staff	8	329	74	23	10.13	16	0.86	0.49	0.84	0.96	0.90	0.3	74.2	4.0
	Outcome	7	582	323	56	32.94	28	0.24	0.80	0.84	0.95	0.90	1.4	39.3	10.5
	Splint	8	432	366	85	52.41	32	0.01	0.82	0.81	0.90	0.85	0.5	13.2	12.5

χ^2 , Chi-square; CA, Cronbach alpha; DF, degrees of freedom; ext, extremes; N, number; PSI, Person Separation Index.

(nontraumatic versus traumatic). The findings provide evidence that each scale's items mapped out a construct that was experienced by most participants in the sample (ie, they scored on the scale). Furthermore, each scale worked well to measure hand function and symptoms for people with nontraumatic and traumatic hand conditions. Table 3 shows other targeting statistics (floor and ceiling effects) as well as missing data, that is, the proportion of eligible participants who left one or more items blank in a scale.

PSI values for the HRQL and treatment outcome scales were 0.80 or greater (see Table 3). PSI values without extremes (≥ 0.77) for the experience of care scales were acceptable and higher than with extremes (≥ 0.49) due to the high ceiling effects for these scales. Cronbach alpha values for the 14 scales were 0.90 or greater (with extremes) to 0.84 or greater (without extremes). The item residual correlations were greater than 0.30 for 10 pairs of items in seven scales. Subtests provided evidence that the correlated items had marginal impact on scale reliability (ie, < 0.05 drop in PSI values). The principal component analysis results for each scale provided broad support for single factors; factor loadings ranged from 0.64 to 0.95, with only six items in three scales (splint, symptoms, and psychological) below 0.70.

Construct validity was examined with the predetermined hypotheses. Table 4 shows the Spearman correlations between HAND-Q scales relating to hypothesis 1. All scales except for outcome correlated more strongly with scales within their domain (eg, HRQL), as hypothesized. The strongest correlations were found in scales that have commonality in the underlying construct, such as function and symptoms, or life impact and psychological.

In regard to hypothesis 2, analysis confirmed that the mean scores were significantly lower ($P < 0.001$ on independent samples *t*-tests) on the appearance scale (62, SD = 24 versus 77, SD = 22) and age-related appearance scale (59, SD = 22 versus 72, SD = 22) for participants with conditions that can affect the appearance of the hands compared with participants with carpal tunnel. The results for hypothesis 3 are graphically demonstrated in Figure 2, which shows that the mean scores by severity of the hand condition were incrementally lower by

increasing severity on the outcome and all HRQL scales ($P \leq 0.002$ on ANOVA), as well as the splint scale ($P = 0.018$). Finally, for hypothesis 4, the mean scores on the HRQL and the outcome scales for the 95 participants in the sample who reported that they need more surgery were lower compared with the 249 who did not need more surgery ($P < 0.022$ on independent samples *t* test). Contrary to our hypothesis, no difference was found on the splint scale score by need for surgery (see Fig. 3). The characteristics of the subgroups for tests of construct validity can be found in figure, Supplemental Digital Content 3, <http://links.lww.com/PRSGO/B896>.

DISCUSSION

The strength of the HAND-Q stems from its robust content validity, which results from the input from a broad cross-section of participants from six countries who have experience of a hand condition and its treatment, as well as expert input from professionals who treat hand conditions.¹⁴ Participants described in detail their experiences of living with a hand condition, the impact their hand condition had on their day-to-day life, and their interactions with the healthcare environment. Based on this rich qualitative data set, we created 20 field-test scales that covered outcomes of interest and the experiences of care. A further strength of our study is the large-scale field test that incorporated multiple languages and enabled vigorous psychometric analysis of the preliminary scales. By using the RUMM2030 software to perform the RMT analysis, we were able to identify that a number of the preliminary scales did not meet stringent expectations for psychometric function. These scales were not included in the final version of the HAND-Q. The remaining 14 scales were found to have strong reliability and validity indicators. These scales measure a broad range of concepts that patients identified as important. The fact that the DIF that was identified did not impact scoring supports international application of the HAND-Q scales for research, clinical practice, and quality improvement applications.

The HAND-Q is a novel hand-specific PROM with a modular structure that provides a comprehensive set of scales, each of which is independently functioning. This

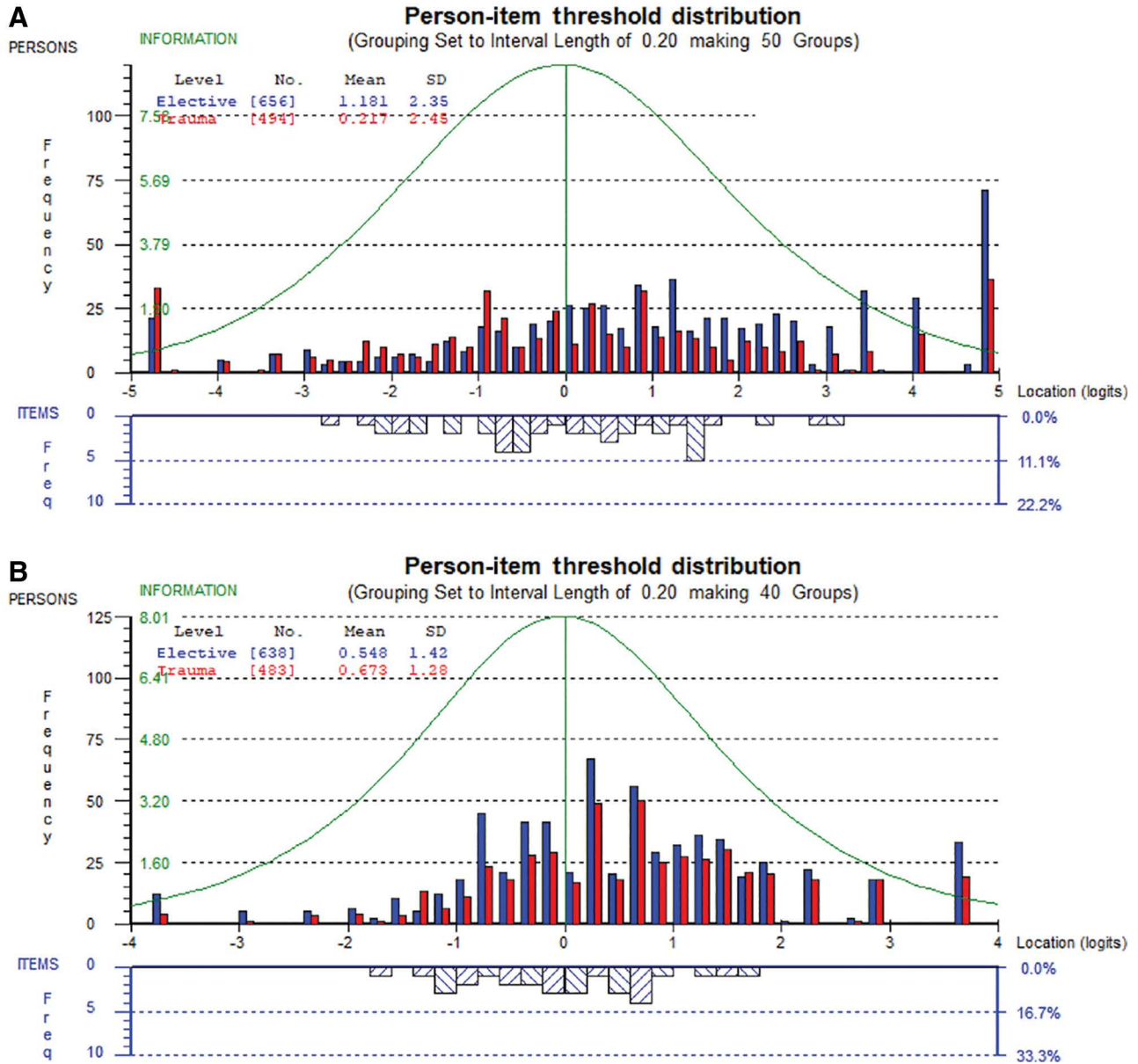


Fig. 1. A, HAND-Q Function scale by type of hand condition. B, HAND-Q Symptom scale by type of hand condition.

Table 4. Spearman Correlations between HAND-Q Scales—Hypothesis 1

	Appearance	Appearance Age-related	Function	Life Impact	Psychological	Sexual	Symptoms	Clinic	Doctor	Hand Therapist	Information	Office Staff	Outcome
Appearance	0.88*												
Appearance age-related													
Function	0.30*	0.28*											
Life impact	0.33*	0.34*	0.62*										
Psychological	0.42*	0.42*	0.50*	0.73*									
Sexual	0.26*	0.23*	0.49*	0.68*	0.55*								
Symptoms	0.41*	0.40*	0.61*	0.65*	0.63*	0.49*							
Clinic	0.23*	0.20*	0.08	0.15*	0.21*	0.17†	0.12†						
Doctor	0.19*	0.17*	0.04	0.16*	0.24*	0.18*	0.16*	0.58*					
Hand therapist	0.23*	0.26*	0.06	0.16†	0.22*	0.19†	0.18*	0.45*	0.54*				
Information	0.33*	0.33*	0.16*	0.33*	0.39*	0.36*	0.30*	0.62*	0.65*	0.51*			
Office staff	0.22*	0.17*	0.13†	0.20*	0.25*	0.29*	0.19*	0.59*	0.55*	0.46*	0.55*		
Outcome	0.34*	0.32*	0.13*	0.22*	0.32*	0.23*	0.28*	0.48*	0.49*	0.45*	0.54*	0.33*	
Splint	0.29*	0.34*	0.17*	0.35*	0.37*	0.33*	0.30*	0.37*	0.31*	0.27*	0.44*	0.25*	0.46*

*Correlation is significant at the 0.01 level (two-tailed).

†Correlation is significant at the 0.05 level (two-tailed).

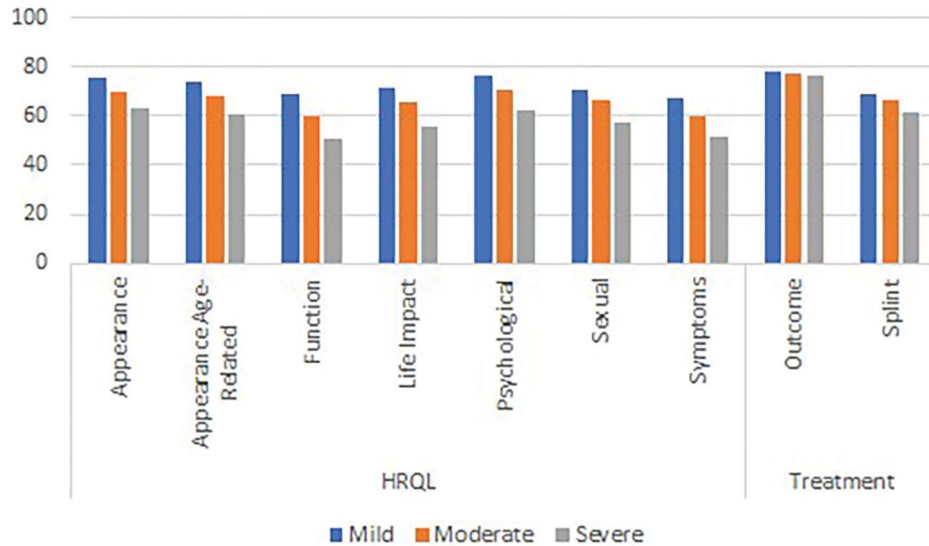


Fig. 2. Mean score for HAND-Q HRQL and splint scales by self-reported condition severity.

attribute allows for tailoring of the scales used to a specific application, thereby minimizing respondent burden by only measuring the concepts of interest. It was observed that the HRQL and treatment outcome scales demonstrated good targeting and distribution of scores, whereas the patient experience scales had larger ceiling effects. Generally, patients report being satisfied with the care that is provided to them and therefore it can be difficult to develop scales that accurately measure patient satisfaction with healthcare services. These experience scales were designed for use in a clinical setting rather than research applications. As such, the ceiling effects are thought to be acceptable; the scales can be used to identify those patients who are not satisfied with the services provided, and to assist with pinpointing the issues to allow for these to be addressed and improved upon.

This study has some limitations. There are numerous conditions that can affect the hands, including congenital, degenerative, and traumatic etiologies. Although our sample has included a wide variety of pathology, it was not possible to evaluate the experience of patients with all types of hand conditions. Rare conditions, such as congenital hand differences and hand transplant, were not included in the sample. Further research into the use of the HAND-Q scales in these conditions would be necessary to support the content validity. Although we exceeded our projected sample of 1000, and were able to include patients from six countries, an important limitation of our sample is that all participants were from high resource countries. The HAND-Q was translated into Urdu, Tamil, Bengali, and Hindi in preparation to include patients from India and Pakistan, but recruitment in these countries was

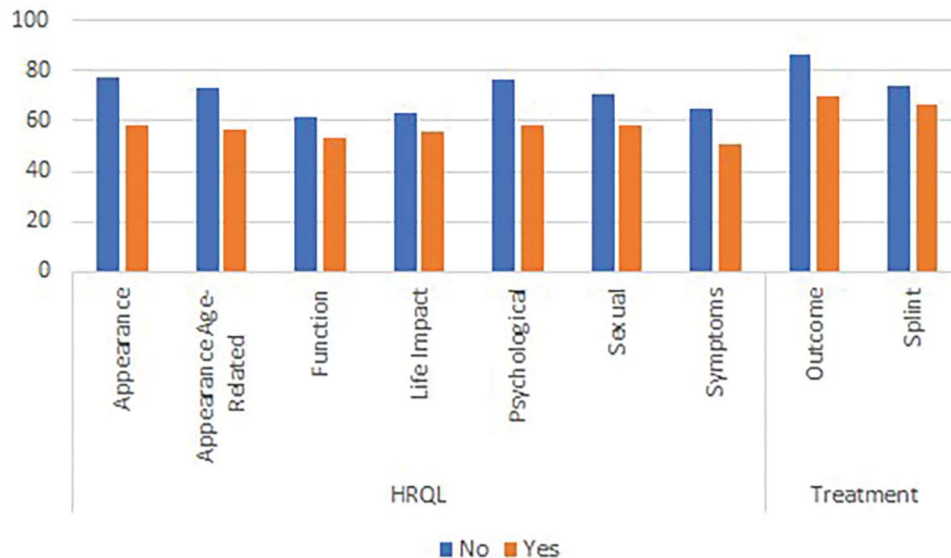


Fig. 3. Mean scores for HAND-Q HRQL and splint scales by need for more surgery.

impacted by the COVID-19 pandemic, and we were not able to include field-test results from these countries. We were, however, able to use feedback about items that were difficult to translate into consideration during the item reduction processes. Future studies to examine the psychometric performance of the HAND-Q in other countries are warranted. Finally, some psychometric properties (responsiveness and test-retest reliability) were not examined in this study and should be examined in future research.

CONCLUSIONS

The HAND-Q was designed for patients with all types of hand conditions, including those presenting for elective surgery and or those associated with trauma. This new modular PROM measures hand function, symptoms, life impact, psychological impact, sexual impact, satisfaction with outcome, and experience of care. This work compliments the existing Q-Portfolio of condition-specific PROMs for surgery patients that includes the BREAST-Q, FACE-Q, BODY-Q, CLEFT-Q, SCAR-Q and WOUND-Q (see www.qportfolio.org). The HAND-Q has been designed from a strong foundation that included extensive involvement of patients and experts in the field, a large multinational field-test and the use of RMT analysis to produce unidimensional scales with interval-level measurement properties. The HAND-Q is a valuable new measurement tool to incorporate the patient perspective into evidence-based clinical care, quality assurance and regulatory decisions in the care of patients with hand conditions.

Kyra L. Sierakowski, MD, PhD

Department of Plastic and Reconstructive Surgery
College of Medicine and Public Health
Flinders University
Adelaide, South Australia
Australia
E-mail: kyra.sierakowski@flinders.edu.au

ACKNOWLEDGMENTS

Development of the HAND-Q has involved hundreds of people with hand conditions and the collaboration of numerous health-care professionals and researchers around the world. We are truly grateful for their dedication and help with our research.

REFERENCES

1. Wormald JCR, Geoghegan L, Sierakowski K, et al. Site-specific patient-reported outcome measures for hand conditions: systematic review of development and psychometric properties. *Plast Reconstr Surg Glob Open*. 2019;7:e2256.
2. Dacombe PJ, Amirfeyz R, Davis T. Patient-reported outcome measures for hand and wrist trauma: is there sufficient evidence of reliability, validity, and responsiveness? *Hand (NY)*. 2016;11:11–21.
3. Lloyd-Hughes H, Geoghegan L, Rodrigues J, et al. Systematic review of the use of patient reported outcome measures in studies of electively-managed hand conditions. *J Hand Surg Asian Pac Vol*. 2019;24:329–341.
4. Sierakowski K, Evans Sanchez, KA, Damarell RA, et al. Measuring quality of life and patient satisfaction in hand conditions: a systematic review of currently available patient reported outcome instruments. *Australas J Plast Surg*. 2018;1:58–99.
5. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med*. 1996;29:602–608.
6. Franchignoni F, Giordano A, Sartorio F, et al. Suggestions for refinement of the Disabilities of the Arm, Shoulder and Hand Outcome Measure (DASH): a factor analysis and Rasch validation study. *Arch Phys Med Rehabil*. 2010;91:1370–1377.
7. Cano SJ, Klassen A, Pusic AL. The science behind quality-of-life measurement: a primer for plastic surgeons. *Plast Reconstr Surg*. 2009;123:98e–106e.
8. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press, 1980.
9. Chang CH, Reeve BB. Item response theory and its applications to patient-reported outcomes measurement. *Eval Health Prof*. 2005;28:264–282.
10. Beaton DE, Wright JG, Katz JN; Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am*. 2005;87:1038–1046.
11. Chung KC, Pillsbury MS, Walters MR, et al. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am*. 1998;23:575–587.
12. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22.
13. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1147–1157.
14. Sierakowski KL, Kaur MN, Sanchez KE, et al. A qualitative study informing the development and content validity of the HAND-Q: a modular patient-reported outcome measure for hand conditions. *BMJ Open* (accepted).
15. Wild D, Grove A, Martin M, et al; ISPOR Task Force for Translation and Cultural Adaptation. 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;8:94–104.
16. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377–381.
17. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess*. 2009;13:iii, ix–ixix, 1.
18. Andrich D. *Rasch Models for Measurement*. Beverley Hills, CA: Sage Publications; 1988.
19. Wright BD, Masters GN. *Rating Scale Analysis*. Chicago: MESA Press; 1982.
20. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–334.
21. Nunnally JC, Bernstein IH. *Psychometric theory*. New York: McGraw-Hill; 1994.
22. Christensen KB, Makransky G, Horton M. Critical values for Yen's Q3: identification of local dependence in the Rasch model using residual correlations. *Appl Psychol Meas*. 2017;41:178–194.
23. Andrich D. An elaboration of Guttman scaling with Rasch models for measurement. *Sociol Methodol*. 1985;15:33–80.
24. Gaskin CJ, Happell B. On exploratory factor analysis: a review of recent evidence, an assessment of current practice, and recommendations for future use. *Int J Nurs Stud*. 2014;51:511–521.
25. Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod*. 2013;38:52–54.