



Rasch analysis of the Disabilities of the Arm, Shoulder and Hand (DASH) instrument in patients with a humeral shaft fracture

Esther M.M. Van Lieshout, MSc, PhD*, Kiran C. Mahabier, MD, PhD, Wim E. Tuinebreijer, MD, PhD, Michael H.J. Verhofstad, MD, PhD, Dennis Den Hartog, MD, PhD, on behalf of the HUMMER Investigators¹

Trauma Research Unit, Department of Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

Background: The Disabilities of the Arm, Shoulder and Hand (DASH) instrument was developed to assess the disability experienced by patients with any musculoskeletal condition of the upper extremity and to monitor change in symptoms and upper-limb function over time. The 30 items are scored on a 5-point rating scale. The Dutch-language version of the DASH instrument (DASH-DLV) has been examined with the classical test theory in patients with a humeral shaft fracture. This study aimed to examine the DASH-DLV with a more rigorous and extensive analysis by applying the Rasch model.

Methods: Data of 400 patients included in a multicenter, prospective study comparing operative and nonoperative treatment of adult patients with a humeral shaft fracture were used. The person-item map, item fit statistics, reliability, response category ordering, and dimensionality were examined. Raw data were converted to linear measures using the Rasch model.

Results: The DASH-DLV showed a good fit to the Rasch model, except for item 26 (“Tingling [pins and needles] in your arm, shoulder or hand”). The person reliability was 0.92. In general, the category functioning of the 5-point rating scale was working well. Dimensionality analysis revealed that the DASH-DLV is a unidimensional scale. Differential item functioning for sex was not detected, and only item 26 exhibited differential item functioning as a function for age.

Conclusion: The DASH-DLV fits the stringent Rasch model in a clinical situation with a group of adult patients with a humeral shaft fracture. Adequate measurement for scientific research can be obtained to evaluate longitudinal intervention research.

Level of evidence: Basic Science Study; Validation of Outcome Instrument

© 2019 Journal of Shoulder and Elbow Surgery Board of Trustees. All rights reserved.

Keywords: PROMs; outcome; humerus; fracture; validity; reliability

This study was exempted by the local Medical Research Ethics Committee of Erasmus MC (no. MEC-2012-296). The medical research ethics committees of all hospitals approved this study.

¹ The HUMMER Investigators are listed at the end of this article.

*Reprint requests: Esther M.M. van Lieshout, MSc, PhD, Trauma Research Unit, Department of Surgery, Erasmus MC, University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, The Netherlands.
E-mail address: e.vanlieshout@erasmusmc.nl (E.M.M. Van Lieshout).

Region-specific patient-reported outcome measures are important instruments for evaluating clinical outcome and functional recovery from the patient's perspective during medical treatment studies. The Disabilities of the Arm, Shoulder and Hand (DASH) instrument is a region-specific patient-reported outcome measure developed in 1996 by a collaborative effort of researchers from the American Academy of Orthopaedic Surgeons and the Institute for Work & Health.¹⁸ It was designed to describe the disability experienced by patients with any musculoskeletal condition of the upper extremity and to monitor change in symptoms and upper-limb function over time.¹⁴ The DASH outcome measure has been validated in over 15 languages in patients with a number of upper-extremity musculoskeletal disorders including rheumatoid arthritis and shoulder impingement syndrome.^{7,18} Normative data have been established for the American and Norwegian populations.^{5,18} In addition, the Dutch-language version of the DASH instrument (DASH-DLV) has been validated in patients with a range of upper-extremity disorders.¹⁵

In a former study, our study group examined the measurement properties of the DASH instrument and other instruments in patients with a humeral shaft fracture by the classical test theory (CTT).¹³ The DASH instrument was shown to be a reliable, valid, responsive, and preferred instrument in this group of patients. After this examination by the CTT, it became appropriate to subject the DASH instrument in patients with a humeral shaft fracture to modern psychometrics based on Rasch measurement theory (RMT). RMT is one of the latent trait models. Latent traits are hypothesized traits or constructs that cannot be directly observed. Items of a rating scale such as the DASH instrument are answered in 5 categories, and the item scores are summed to provide a total score, which provides only an ordinal level of measurement. Ordinal data are not equal-interval data and are not linear data. Despite frequent use, ordinal data should not be evaluated using parametric analyses, such as *t* tests and analysis of variance. In the CTT, the total score is the summation of the person's true score and error score. This equation contains no formalization of the difficulty of an item. RMT is a probabilistic module: The probability of scoring a certain item category is dependent on the difficulty of the item and the disability of the person. We chose RMT because it is considered superior to the CTT as it makes stronger assumptions and provides stronger findings. For this reason, RMT is nowadays frequently applied in quality-of-life research.⁶ Using RMT modeling involves a rigorous and extensive analysis of the data and provides additional psychometric information that cannot be obtained through the CTT approach. The Rasch model thus provides a much more inclusive assessment of overall fit and appropriateness. The data are tested for fit into the Rasch model, allowing for a detailed examination of the internal construct validity of the scale, including properties such as reliability and ordering of the

categories. The Rasch model also determines whether a scale is unidimensional, which is required to justify summation of scores and can linearly transform raw scores from their original scale to an equal-interval scale to allow application of parametric statistics. In particular, the use of sum scores for longitudinal multi-item questionnaire data can lead to biased parameter estimates, which can be prevented by the use of RMT-based variable scores.⁴ As we have performed a longitudinal study in patients who have sustained a humeral shaft fracture,¹² a data set could be generated to enable the application of the Rasch model to the DASH-DLV. The suitability of the DASH instrument also depends on the cause of the disability of the arm, shoulder, and hand.¹ This implies that our results are not generalizable to other extremity conditions.

The aims of this study were therefore to estimate linear measures for disability in patients with a humeral shaft fracture who completed the DASH-DLV at 5 time points, to assess dimensionality as a type of construct validity of the DASH-DLV, and to evaluate the presence of differential item functioning (DIF) of the DASH-DLV.

Materials and methods

The data of the first 400 patients included in a multicenter, prospective cohort study comparing operative and nonoperative treatment of adult patients with a humeral shaft fracture were used. For the current study, no comparison between operative and nonoperative treatment is considered. This study is registered with the Netherlands Trial Register (no. NTR3617). The study protocol for this study has been published elsewhere.¹² All patients provided signed informed consent.

Study population

Patients aged 18 years or older presenting with a humeral shaft fracture (AO type 12A or 12B) to the emergency department of 1 of 32 participating hospitals in the Netherlands were included. The exclusion criteria were concomitant injuries affecting treatment and rehabilitation of the affected arm; treatment with an external fixator; pathologic, recurrent, or open fractures; neurovascular injuries requiring immediate surgery (excluding radial nerve palsy); additional traumatic injuries of the affected arm that influenced upper-extremity function; impaired upper-extremity function prior to the injury; retained hardware around the affected humerus prior to trauma; rheumatoid arthritis; any bone disorder possibly impairing bone healing (excluding osteoporosis); problems with ensuring follow-up (eg, cognitive impairment or no fixed address); or insufficient comprehension of the Dutch language.

DASH instrument

Patients were asked to complete the DASH-DLV.¹³ The DASH instrument is easy to use and can be completed by any patient, regardless of age. The DASH questionnaire was developed to

describe disabilities experienced by patients with any musculo-skeletal condition of the upper extremity and to monitor change in symptoms and upper-limb function over time.⁵ The DASH questionnaire contains 2 optional 4-item modules enabling measurement of symptoms and upper-extremity dysfunction in athletes, performing artists, and other workers whose jobs require more advanced physical activity. The DASH questionnaire is scored in 2 components: the disability/symptom items (30 items, scored 1-5) and the optional modules (4 items, scored 1-5). The DASH score is calculated using the following formula: $((\text{Sum of all items/No. of questions answered}) - 1) \times 25$. Alternatively, an analogous formula can be used in case up to 3 items are missing. The overall score ranges from 0 to 100 points. The average is taken of all obtained values for completed responses, producing a score out of 5 points, which is transformed to a score out of 100 points. High scores represent higher disability. Patients needed to have completed at least 27 of the 30 disability/symptom items of the DASH questionnaire to enable calculation of the total DASH score.⁸ The 2 optional 4-item modules did not apply to the current study population and were therefore not used.

Statistical analysis

The DASH data were transferred into the Rasch rating scale model using Winsteps measurement software (Winsteps Rasch-model computer programs; Winsteps.com, Beaverton, OR, USA). A rating scale parameterization was used because all the items have the same number of categories. The following analyses were performed:

1. Construction of the person-item map (Wright map)
2. Testing of (mis)fit between the data and the model
3. Estimation of the person and item reliability and separation coefficient
4. Testing of the ordering of the categories
5. Analysis of dimensionality
6. Examination of local dependency
7. Evaluation of DIF
8. Conversion of the logit scale to more meaningful units

Person-item, or Wright, map

A map was constructed of the hierarchy for the person and item measures for the DASH instrument to examine person and item performances. A person measure is a quantitative measure of a person's disability of the upper extremity on a unidimensional scale. An item measure is a quantitative measure of the item's difficulty of accomplished activities or appearance of symptoms caused by disabilities of the upper extremity. Person and item measures are expressed in the same units of logits. At the bottom of the map, the lower estimates of the person and item can be found, with increasing estimates represented higher up the map. On the left side, the patient performances are represented, and on the right side, the items. For a well-targeted measure, the mean location for the person would be around 0 logits.

Testing of (mis)fit between data and model (item fit statistics)

To determine how well the empirical data fit the Rasch model, χ^2 fit statistics are calculated. These fit statistics are the infit mean

square (infit MNSQ) and the outfit mean square (outfit MNSQ). The infit mean square represents the information-weighted mean square residual difference between observed and expected responses. The infit statistics are sensitive to unexpected responses near the person's ability level. The outfit statistic is the usual unweighted mean square residual and is more sensitive to outliers. The expected infit or outfit mean square value is 1.0. A mean square value greater than 2.0 indicates more misinformation than information. Values should range between 0.5 and 1.7 for clinical observations.¹⁶ High infit and outfit reflect underfit, which means lack of predictability of an item. Low infit and outfit reflect overfit, which means over-predictability of an item.

Reliability and separation statistics

In the Rasch model, reliability is estimated both for persons and for items. Person reliability in Winsteps is equivalent to the test reliability (Cronbach α) in the CTT. The person reliability reports how reproducible the person's ability order is in this sample of persons for this set of items. The item reliability reports how reproducible the item's difficulty order is for this set of items for this sample of persons. The higher the separation, the better the instrument is in differentiating person ability and item difficulty. Separation is measured on a continuous scale bounded by 0 and infinity, which is an advantage over psychometric reliability, which only ranges from 0 to 1. The person separation index can be used to calculate the number of distinct levels of scar quality (strata) that the items can distinguish ($\text{Strata} = [4 \times \text{Person separation index} + 1]/3$).^{3,17}

Category function

Category functioning is examined by analyzing category frequencies, average measures, thresholds, and category fit statistics.⁸ The items of the DASH instrument have 5 categories. The category frequencies indicate how many patients choose a particular response category. The recommended minimal number of responses per category is 10 for stable rating-scale-structure threshold parameter estimates. The average measures are defined as the average of the ability estimates for all patients in the sample, which are assigned by the patients to that particular response category, with the average calculated across all the observations in that category.¹⁷ The average measures and the thresholds should increase when moving from lower to higher categories. Guidelines recommend that thresholds should increase by at least 1.4 logits to show distinction between categories—but not by more than 5 logits. In item characteristic curves, the probabilities of choosing a certain category are plotted against the latent variable. When there are ordered categories, the category probability curves show that each category is the most probable category at some point on the latent variable. Fit statistics provide another criterion for assessing the quality of the rating scale. Outfit mean squares greater than 2 indicate more misinformation than information. The category has been used unexpectedly, and there is unexplained noise.

Dimensionality investigation

According to the Rasch methodology, when the data fit the Rasch model, the Rasch dimension is the only dimension in the data. Rasch factor analysis is a factor analysis of the residuals that remain after the linear Rasch measure has been extracted from the data set. A secondary dimension in the data must explain at least 2 items' worth of

variance: Unless a component has the strength of at least 2 items, it may merely be due to an idiosyncratic item. The residual variance or unexplained variance is expressed by eigenvalue units. The eigenvalue implies the number of items that are off dimension.¹⁰

Local dependency

Local independence is a central assumption of the RMT model. Response dependency is a form of local dependency. Response dependency is the linking of items by making the response to 1 item determine the response to another. An example is when 2 stair-climbing items are included in the same scale. If one can climb several flights of stairs unaided, one must be able to climb 1 flight of stairs. This results in biased parameters and inflating reliability. Local dependency among items was examined by identifying correlations among the residuals of the items. Correlated item residuals 0.3 above the average of all item residuals are considered locally dependent.²

Comparing sex for DIF

In the Rasch model, the estimated measures should be invariant across groups, such as female and male patients. The hierarchy of the items is assumed to be the same across groups: It should work uniformly, irrespective of groups. The size of the item measures will be compared graphically across groups. DIF means that female disability and male disability differ across separate items. The DIF measure is the absolute difficulty of the item for the group. The difference between the 2 DIF measures is called the “DIF contrast.” A DIF contrast greater than 0.64 logits has been used for detecting DIF.¹³ Another way to detect DIF is to look at statistically significant ($P < .05$) Rasch-Welch t tests and Mantel tests. More than 1 item on a scale, or more than 5% of the items, should demonstrate DIF to distinguish disability between groups of patients.⁹

Converting logit scale to more meaningful units

The item measures in logits were rescaled to another equal-interval, user-friendly Rasch scale with a range of 0 to 100.

Results

The data collection resulted in the use of 1773 DASH scores taken from 400 patients who filled out the questionnaire at 2 and 6 weeks and at 3, 6, and 12 months after trauma. Data from all time points were used to have a broad range in scores, from low to high disability. Overall, 165 patients (41%) were men, and 252 patients (63%) were treated operatively. The median age was 58 years (P_{25} - P_{75} , 40-69). The injured upper limb was the dominant limb in 195 patients (49%). In the Rasch analysis, 1638 DASH measures were included because 135 questionnaires had extreme scores (eg, putting a cross against category 1 or 5 for all items). These extreme scores do not provide useful information for comparing item difficulties.

Person-item, or Wright, map

Figure 1 presents the person-item map. The items on the right side are displayed along the logit scale in the order of

measurement. The default mean difficulty is set at 0 and has a standard deviation of 0.57. The DASH items cover 2.45 logits (range, -1.03 to 1.45). The items at the lowest level are those that the patients easily endorsed, for example, item 18 (“Recreational activities which require you to take some force or impact through your arm, shoulder or hand [eg, golf, hammering, tennis, etc]”) and item 30 (“I feel less capable, less confident or less useful because of my arm, shoulder or hand problem”). The items at the top are those that the patients found difficult to endorse, for example, item 3 (“Turn a key”).

Patients are plotted at the left side. Most patients are located at the middle of the map opposite the items or below the middle of the map below the items. The first group is well targeted, meaning that the items represent the patients’ level of disability. The last patient group is only lightly endorsed after the humeral shaft fracture and has few disabilities because they are healed 6 or 12 months after the fracture. The mean patient ability measure is -1.29 logits (standard deviation, 1.64 logits), which is more than 1 logit below the average difficulty of the items (ie, the local origin, which is set at 0).

Item statistics table

Table I shows the items of the DASH instrument placed according to the hierarchy of the item difficulties. The measures are the item difficulty estimates in logits. Except for items 3 and 26, 21 and 28, 23 and 8, and 8 and 19, the inter-item separation of the items is less than 0.15 logits, indicating overlap between items. All the items, except item 26 (“Tingling [pins and needles] in your arm, shoulder or hand”), have mean square infit or outfit values between 0.5 and 1.7. Item 26 (“Tingling”) has an outfit mean square of 2.59.

Reliability and separation statistics

Reliability analyses revealed that the item reliability coefficient of the DASH instrument was 1.00 and the item separation coefficient was 19.26. The person reliability was 0.92, and the person separation was 3.40. The person separation index was used to calculate the number of distinct levels of disability (strata) that can be differentiated. Five ranges of disability can be confidentially distinguished.

Category functioning

Table II presents the functioning of the 5 categories of the items (5-point Likert scale). Category 1 is the most represented category, with a frequency of 26,189 observations. This one category included the patients with the least disability. The observed average measures advance monotonically in a smooth distribution from -1.88 to 0.63 logits.

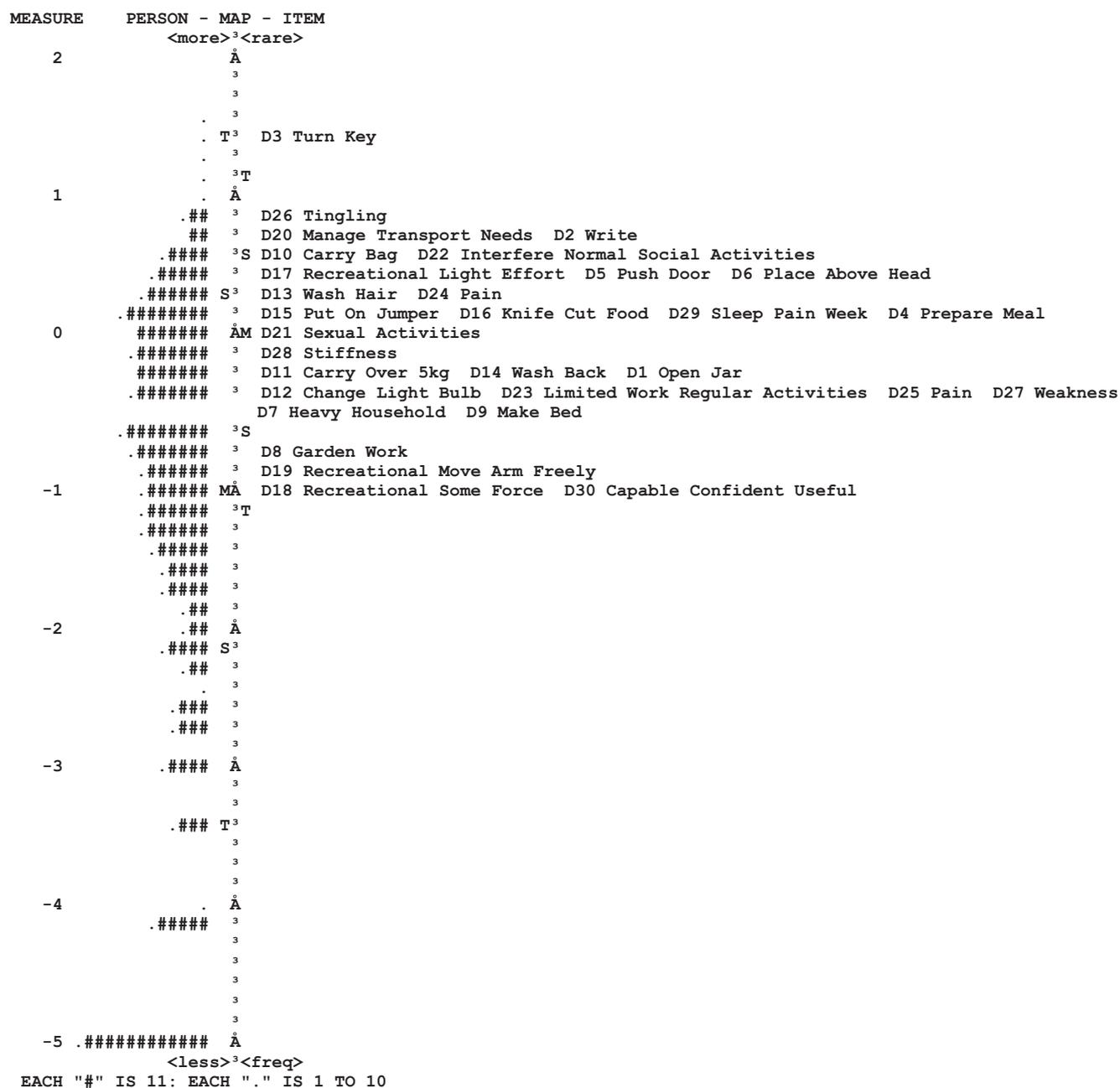


Figure 1 Person ($n = 355$) and item ($n = 30$) or Wright map for Disabilities of the Arm, Shoulder and Hand instrument. Positive scores (in logits) indicate poorer abilities, whereas negative scores demonstrate better abilities of the arm, shoulder, and hand. Items from the scale are shown on the right side of the figure, and person measures are highlighted by a *dot* or *#*. Each *dot* represents 1 to 10 subjects, and each *#* represents 11 subjects. *M*, mean; *S*, 1 standard deviation from mean; *T*, 2 standard deviations from mean.

The thresholds of categories 2, 3, and 4 increase monotonically. The threshold of category 5 does not increase. None of the thresholds increases at least 1.4 logits. None of the categories shows a misfit.

Figure 2 shows the category probability curves of the DASH categories with a smooth distribution. Thresholds are ordered. Only the threshold between the fourth and fifth categories is unclear. In this Rasch-Andrich model (1 of the polytomous models), the rating scale structure is defined to

be equal for all items. The category rating scale is working well.

Local dependency

Local dependency was examined by identifying correlations among the residuals of the items. The average of all item residual correlations was -0.030 . Table III

Table I Item statistics of DASH instrument

DASH item	Count	Measure	Infit MNSQ	Outfit MNSQ
3: Turn key	1769	1.45	1.23	0.80
26: Tingling	1764	0.84	1.48	2.59
20: Manage transport needs	1766	0.75	1.25	0.99
2: Write	1769	0.75	1.34	0.99
22: Interfere normal social activities	1764	0.62	0.93	0.87
10: Carry bag	1772	0.55	1.22	1.06
5: Push door	1765	0.45	1.10	1.08
17: Recreational light effort	1758	0.40	1.11	0.79
6: Place above head	1767	0.37	1.49	1.43
13: Wash hair	1771	0.27	1.32	0.70
24: Pain	1767	0.24	0.84	0.95
15: Put on jumper	1766	0.14	0.84	0.68
4: Prepare meal	1772	0.10	0.77	0.59
16: Knife cut food	1768	0.08	0.85	0.66
29: Sleep pain week	1766	0.08	0.99	1.11
21: Sexual activities	1697	0.05	1.01	1.31
28: Stiffness	1765	-0.15	1.21	1.54
11: Carry over 5 kg	1766	-0.23	1.15	1.10
1: Open jar	1772	-0.24	1.01	0.92
14: Wash back	1772	-0.35	0.95	0.87
12: Change light bulb	1771	-0.38	1.26	1.10
9: Make bed	1763	-0.40	0.81	0.73
27: Weakness	1763	-0.40	1.07	1.31
25: Pain specific activity	1763	-0.43	0.77	1.21
7: Heavy household	1770	-0.45	0.76	0.68
23: Limited work regular activities	1767	-0.47	0.74	0.76
8: Garden work	1755	-0.72	0.73	0.69
19: Recreational move arm freely	1771	-0.90	1.08	1.02
30: Capable confident useful	1766	-0.95	1.17	1.14
18: Recreational some force	1770	-1.03	0.93	0.85

DASH, Disabilities of the Arm, Shoulder and Hand; *MNSQ*, mean square.

The items are listed according to the hierarchy of the item difficulties ("Measure"). The higher items are concerned with the highest disability of the arm, shoulder, and hand. Infit or outfit MNSQ values have a reasonable range of 0.5 to 1.7.

presents the items with a residual item correlation more than 0.3 higher than the average of all item residual correlations.

Dimensionality investigation

The raw variance of the DASH instrument explained by Rasch measures is 43.8% (expected by model, 43.4%). The unexplained variance in the first contrast is 5.2% (3.6 eigenvalue units). The first contrast consists of the symptoms "Arm, shoulder or hand pain," "Stiffness in your arm, shoulder or hand," "Weakness in your arm, shoulder or hand," and "Arm, shoulder or hand pain when you do any specific activity" vs. the activities "Prepare a meal," "Garden or outdoor property work," "Do heavy household jobs (eg, wash windows, clean floors)," and "Use a knife to cut food."

Differential item functioning

None of the items showed DIF contrast measures for sex greater than 0.65 logits. Only 1 item ("Tingling [pins and

needles] in your arm, shoulder or hand") revealed a DIF contrast measure of 0.69 logits for patients aged 59 years or older.

Converting logit scale to more meaningful units (user-friendly rescaling)

The range of the Rasch person measures in logits was transformed to the range of 1 to 100 (Table IV). The formula for predicting the rescaled measure from the logits is as follows: Measure = Logit measure \times 9.7961 + 51.6042.

Discussion

To our knowledge, no other study has analyzed the DASH instrument using the Rasch model. Modern test theory analysis on the disability assessment scale of the DASH instrument is important to improve the evidence base in humeral shaft fracture treatment. The DASH instrument performed well in a group of patients with a humeral shaft

Table II Summary of category structure of DASH instrument

Category label/score	Observed count	Observed count %	Observed average	Outfit MNSQ	Threshold
1	26,189	49	-1.88	1.10	None
2	9912	19	-0.87	0.94	-0.50
3	7123	13	-0.23	0.91	-0.18
4	4145	8	0.23	1.04	0.55
5	5566	11	0.63	1.11	0.14

DASH, Disabilities of the Arm, Shoulder and Hand; *MNSQ*, mean square. Outfit MNSQ values have a reasonable range of 0.5 to 1.7.

fracture. The person reliability is well above 0.8, which is the lower limit of reliability required for serious decision making.¹⁰ Meaningful reliability or reproducibility of the measure is to make decisions in clinical medicine. Person reliability in RMT is comparable with the Cronbach α in the CTT, and it is bounded by 0 to 1. For making decisions in individual patients, one needs higher reliability than for making decisions at the group level, for instance, comparing 2 treatment groups. The item reliability for this sample of patients is very good. Five statistically distinct levels of disability can be differentiated, that is, extreme, severe, moderate, mild, and no disability.

The items of the DASH instrument are intended to measure a single or unidimensional variable, being

disability of the arm, shoulder, and hand. No substantial dimension could be identified by Rasch factor analysis, indicating that the DASH instrument is a suitable unidimensional questionnaire for patients with a humeral shaft fracture. This finding is in accordance with CTT principal component factor analysis performed by Veehof et al¹⁵ examining the DASH-DLV.¹³ However, the dimensionality investigation of the DASH instrument shows an interesting structure: The items “Stiffness in your arm, shoulder or hand” and “Weakness in your arm, shoulder or hand” can be interpreted as a sub-dimension of neurologic symptoms in disability. This structure also reveals a contrast between items for activities vs. items for symptoms. In future studies comparing treatment modalities, it

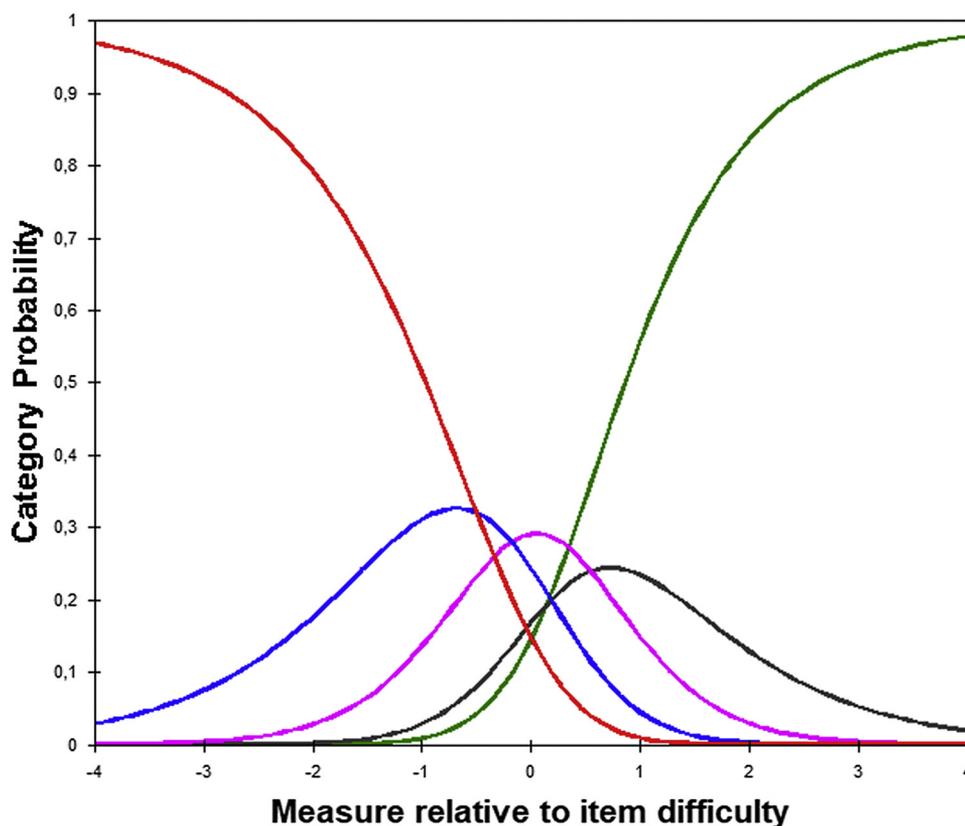


Figure 2 Category probability curve of the Disabilities of the Arm, Shoulder and Hand instrument showing the probability of being assigned to any particular category (*y-axis*), given the difference in estimates between any patient disability and any item difficulty. The threshold estimates correspond to the intersection of rating scale categories.

Table III Local dependent items with their residual item correlations

Local dependent items	Residual item correlation
5: Push heavy door/6: Place above head	0.278
4: Prepare a meal/16: Use knife to cut food	0.285
27: Weakness/28: Stiffness	0.370
7: Heavy household job/8: Garden work	0.399
10: Carry bag/11: Carry over 5 kg	0.411
18: Recreational some force/19: Recreational move arm freely	0.499
24: Pain/25: Pain specific activity	0.504

Items with a residual item correlation 0.3 higher than the average correlation are presented.

would be interesting to compare mean sum scores for activities vs. for symptoms by calculating sum scores for the first 23 items (activities) and for the last 7 items (symptoms).

The items of the Wright map of the DASH instrument show that items 3 (“Turn a key”) and 26 (“Tingling [pins and needles] in your arm, shoulder or hand”) have a high item difficulty without overlap, meaning that the patients assess these activities as the most severe in relation to their arm, shoulder, and hand disability. Item 18 (“Recreational activities which require you to take some force or impact through your arm, shoulder or hand [eg, golf, hammering, tennis, etc]”) and item 30 (“I feel less capable, less confident or less useful because of my arm, shoulder or hand problem”) have a low item difficulty with some overlap, meaning that the patients assess this activity and

this feeling as the least hard in relation to their arm, shoulder, and hand disability.

Table II shows that the category frequencies of the items are highly skewed to the lower end. The distribution of patient measures in Figure 1 is also skewed to the lower end. This skewness is caused by the fact that many patients are achieving full recovery and report no disabilities at some time point after trauma.

The DASH fit statistics revealed a good fit for clinical observations. Item 26 (“Tingling [pins and needles] in your arm, shoulder or hand”), with a mean square outfit measure of 2.59 and infit measure of 1.48 logits, had the highest item fit statistic, reflecting some unpredictability (ie, erratic [unreliable, unpredictable] responses or noise). Item 4 (“Prepare a meal”), with a mean square outfit measure of 0.59 and infit measure of 0.77 logits, had a low item fit

Table IV Raw DASH scores (from 30 to 150) with Rasch logits converted to measures from 0 to 100

Score	Measure										
30	0.00	51	40.13	72	47.25	93	52.62	114	58.11	135	66.07
31	11.67	52	40.58	73	47.52	94	52.87	115	58.40	136	66.64
32	18.31	53	41.00	74	47.80	95	53.12	116	58.70	137	67.25
33	22.14	54	41.41	75	48.06	96	53.37	117	59.00	138	67.91
34	24.84	55	41.81	76	48.33	97	53.61	118	59.31	139	68.61
35	26.91	56	42.19	77	48.59	98	53.87	119	59.62	140	69.38
36	28.60	57	42.56	78	48.85	99	54.12	120	59.94	141	70.23
37	30.02	58	42.93	79	49.11	100	54.37	121	60.26	142	71.17
38	31.24	59	43.28	80	49.37	101	54.62	122	60.60	143	72.25
39	32.32	60	43.62	81	49.63	102	54.88	123	60.94	144	73.48
40	33.28	61	43.96	82	49.88	103	55.13	124	61.29	145	74.96
41	34.16	62	44.28	83	50.13	104	55.39	125	61.65	146	76.78
42	34.95	63	44.60	84	50.38	105	55.65	126	62.02	147	79.16
43	35.68	64	44.92	85	50.63	106	55.91	127	62.40	148	82.61
44	36.36	65	45.23	86	50.88	107	56.18	128	62.80	149	88.76
45	37.00	66	45.53	87	51.13	108	56.45	129	63.21	150	100.00
46	37.59	67	45.83	88	51.38	109	56.71	130	63.63		
47	38.15	68	46.12	89	51.63	110	56.99	131	64.07		
48	38.69	69	46.41	90	51.88	111	57.26	132	64.54		
49	39.19	70	46.69	91	52.13	112	57.54	133	65.02		
50	39.67	71	46.97	92	52.37	113	57.83	134	65.53		

DASH, Disabilities of the Arm, Shoulder and Hand.

statistic, indicating that this item has too much predictability: There is less variation in the data than in the model.

The category rating scale of the DASH instrument is working well, although the fourth category is masked by categories 3 and 5 in the category probability curves. The underuse of the high categories in our population can cause disordering of thresholds.¹¹

Some items showed potential problematic local dependency because these items reflect activities or symptoms that are linked in some way, such that the response on 1 item governs the response on another item because of similarities in item content or response format. An example is item 4 (“Prepare a meal”) and item 16 (“Use a knife to cut food”). If a person can prepare a meal, then he or she must be able to cut food. Another example is item 25 (“Arm, shoulder or hand pain when you do any specific activity”) and item 24 (“Arm, shoulder or hand pain”). If a person has arm, shoulder, or hand pain, then he or she can have arm, shoulder, or hand pain when doing any specific activity. The problems with violations of local dependency are the influence on estimation of person parameters and inflated estimates of reliability.

The hierarchy of the items is assumed to be the same across groups of patients. DIF was not detected for sex and was detected for only 1 item for age of 59 years or older vs. age younger than 59 years. This means that DASH sum scores or scores of the items in male and female patients or in older and younger patients can simply be compared.

Rasch person measures are in comparison with raw data of the rating scale DASH linear data. Especially strong nonlinearity is observed for extremely low raw scores, which develops in longitudinal studies when patients are improving.⁵ For this reason, we present [Table III](#) for converting raw DASH scores to linear measures in a user-friendly way prior to analysis with parametric statistical tests.

Conclusion

This study revealed several valuable insights on the psychometric properties of the DASH-DLV. The analyses confirmed that the scale is reliable and provides a unidimensional measure for disability of the arm, shoulder, and hand. Local dependency of items could have inflated the reliability. For patients with a humeral shaft fracture, all items except item 26 (“Tingling [pins and needles] in your arm, shoulder or hand”) showed a good fit to the stringent Rasch model. We believe that the functioning of this item is highly dependent on the occurrence of radial nerve palsy. The category functioning should be examined in a group of patients with more patient scores in the middle of the scale before it is concluded that 4 categories instead of 5 categories are enough for patients with a humeral shaft fracture. For 2

distinct groups of sex and age, the items do not have significantly different meanings.

The DASH-DLV fits the stringent Rasch model in a clinical situation with a group of adult patients with a humeral shaft fracture. Adequate measurement for scientific research can be obtained to evaluate longitudinal intervention research.

Disclaimer

This work was funded by a grant from the Osteosynthesis and Trauma Care Foundation (reference no. 2013-DHEL).

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

HUMMER Investigators

The HUMMER Investigators are as follows: The local principal investigators comprise Hugo W. Bolhuis, MD, Department of Surgery, Gelre Hospital, Apeldoorn, The Netherlands; P. Koen Bos, MD, PhD, Department of Orthopaedic Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands; Maarten W.G.A. Bronkhorst, MD, PhD, Department of Surgery, Haaglanden MC, The Hague, The Netherlands; Milko M.M. Bruijninx, MD, Department of Surgery, IJsselland Hospital, Capelle aan den IJssel, The Netherlands; Jeroen De Haan, MD, PhD, Department of Surgery, Dijklanderziekenhuis, Hoorn, The Netherlands; Axel R. Deenik, MD, PhD, Department of Orthopaedic Surgery, Haaglanden MC, The Hague, The Netherlands; P. Ted Den Hoed, MD, PhD, Department of Surgery, Ikazia Hospital, Rotterdam, The Netherlands; Martin G. Eversdijk, MD, Department of Surgery, St. Jansdal Hospital, Harderwijk, The Netherlands; J. Carel Goslings, MD, PhD, Trauma Unit Department of Surgery, Academic Medical Center, Amsterdam, The Netherlands; Robert Haverlag, MD, Department of Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands; Martin J. Heetveld, MD, PhD, Department of Surgery, Spaarne Gasthuis, Haarlem, The Netherlands; Albertus J.H. Kerver, MD, PhD, Department of Surgery, Franciscus Gasthuis & Vlietland, Rotterdam, The Netherlands; Karel A. Kolkman, MD, Department of Surgery, Rijnstate, Arnhem, The Netherlands; Peter A. Leenhouts, MD, Department of Surgery, Zaans Medical Center, Zaandam, The Netherlands; Sven A.G. Meylaerts, MD, PhD, Department of Surgery, Haaglanden MC, The

Hague, The Netherlands; Ron Onstenk, MD, Department of Orthopaedic Surgery, Groene Hart Hospital, Gouda, The Netherlands; Martijn Poeze, MD, PhD, Department of Trauma Surgery, Maastricht University Medical Center, Maastricht, The Netherlands; Rudolf W. Poolman, MD, PhD, Department of Orthopaedic Surgery, OLVG, Amsterdam, The Netherlands; Bas J. Punt, MD, Department of Surgery, Albert Schweitzer Hospital, Dordrecht, The Netherlands; Ewan D. Ritchie, MD, Department of Surgery, Alrijne Hospital, Leiderdorp, The Netherlands; W. Herbert Roerdink, MD, PhD, Department of Surgery, Deventer Hospital, Deventer, The Netherlands; Gert R. Roukema, MD, Department of Surgery, Maasstad Hospital, Rotterdam, The Netherlands; Jan Bernard Sintenie, MD, Department of Surgery, Elkerliek Hospital, Helmond, The Netherlands; Nicolaj M.R. Soesman, MD, Department of Surgery, Franciscus Gasthuis & Vlietland, Schiedam, The Netherlands; Edgar J.T. Ten Holder, MD, Department of Orthopaedic Surgery, IJsselland Hospital, Capelle aan den IJssel, The Netherlands; Maarten Van der Elst, MD, PhD, Department of Surgery, Reinier de Graaf Gasthuis, Delft, The Netherlands; Frank H.W.M. Van der Heijden, MD, PhD, Department of Surgery, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands; Frits M. Van der Linden, MD, Department of Surgery, Groene Hart Hospital, Gouda, The Netherlands; Peer Van der Zwaal, MD, PhD, Department of Orthopaedic Surgery, Haaglanden MC, The Hague, The Netherlands; Jan P. Van Dijk, MD, Department of Surgery, Hospital Gelderse Vallei, Ede, The Netherlands; Hans-Peter W. Van Jonbergen, MD, PhD, Department of Orthopaedic Surgery, Deventer Hospital, The Netherlands; Egbert J.M.M. Verleisdonk, MD, PhD, Department of Surgery, Diaconessenhuis, Utrecht, The Netherlands; Jos P.A.M. Vroemen, MD, PhD, Department of Surgery, Amphia Hospital, Breda, The Netherlands; Marco Waleboer, MD, Department of Surgery, Admiraal De Ruyter Hospital, Goes, The Netherlands; Philippe Wittich, MD, PhD, Department of Surgery, St. Antonius Hospital, Nieuwegein, The Netherlands; and Wietse P. Zuidema, MD, Department of Trauma Surgery, VU University Medical Center, Amsterdam, The Netherlands. The medical students (Trauma Research Unit, Department of Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands) comprise Ahmed Al Khanim, Jelle E. Bou-sema, Kevin Cheng, Yordy Claes, J. Daniël Clossen, Emmelie N. Dekker, Aron J.M. De Zwart, Priscilla A. Jawahier, Boudijn S.H. Joling, Cornelia (Marije) A.W. Notenboom, Jaap B. Schulte, Nina Theyskens, Gijs J.J. Van Aert, Saskia H. Van Bergen, Boyd C.P. Van der Schaaf, Tim Van der Torre, Joyce Van Veldhuizen, Lois M.M. Verhagen, Maarten Verwer, and Joris Vollbrandt.

References

1. Baker K, Barrett L, Playford ED, Aspden T, Riazi A, Hobart J. Measuring arm function early after stroke: is the DASH good enough? *Neurol Neurosurg Psychiatry* 2016;87:604-10. <https://doi.org/10.1136/jnnp-2015-310557>
2. 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;41:178-94. <https://doi.org/10.1177/0146621616677520>
3. Fisher WP Jr. Reliability statistics. In: Rasch Measurement Transactions; 1992. p. 238. Available at: <http://www.rasch.org/rmt/rmt63i.htm>. Accessed April 04, 2019.
4. Gorter R, Fox J, Twisk JW. Why item response theory should be used for longitudinal questionnaire data analysis in medical research. *BMC Med Res Methodol* 2015;15:55. <https://doi.org/10.1186/s12874-015-0050-x>
5. Haldorsen B, Svege I, Roe Y, Bergland A. Reliability and validity of the Norwegian version of the Disabilities of the Arm, Shoulder and Hand questionnaire in patients with shoulder impingement syndrome. *BMC Musculoskelet Disord* 2014;15:78. <https://doi.org/10.1186/1471-2474-15-78>
6. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care* 2000;38(9 Suppl):I28-42.
7. 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;29:602-8.
8. Kennedy CA, Beaton DE, Solway S, McConnell S, Bombardier C. Disabilities of the Arm, Shoulder and Hand (DASH). The DASH and QuickDASH Outcome Measure User's Manual. 3rd ed. Toronto, Ontario: Institute for Work & Health; 2011.
9. Lerdal A, Kottorp A. Psychometric properties of the Fatigue Severity Scale—Rasch analyses of individual responses in a Norwegian stroke cohort. *Int J Nurs Stud* 2011;48:1258-65. <https://doi.org/10.1016/j.ijnurstu.2011.02.019>
10. Linacre JM. A user's guide to WINSTEPS: Rasch-model computer programs. *Winsteps.com*; 2019.
11. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002;3:85-106.
12. Mahabier KC, Den Hartog D, Theyskens N, Verhofstad MHJ, Van Lieshout EMM, HUMMER Trial Investigators. Reliability, validity, responsiveness, and minimal important change of the Disabilities of the Arm, Shoulder and Hand and Constant-Murley scores in patients with a humeral shaft fracture. *J Shoulder Elbow Surg* 2017;26:e1-12. <https://doi.org/10.1016/j.jse.2016.07.072>
13. Mahabier KC, Van Lieshout EMM, Bolhuis HW, Bos PK, Bronkhorst MWGA, Bruijninx M, et al. HUMeral Shaft Fractures: MEasuring Recovery after Operative versus Non-operative Treatment (HUMMER): a multicenter comparative observational study. *BMC Musculoskelet Disord* 2014;15:39. <https://doi.org/10.1186/1471-2474-15-39>
14. Slobogean GP, Noonan VK, O'Brien PJ. The reliability and validity of the Disabilities of Arm, Shoulder, and Hand, EuroQol-5D, Health Utilities Index, and Short Form-6D outcome instruments in patients with proximal humeral fractures. *J Shoulder Elbow Surg* 2010;19:342-8. <https://doi.org/10.1016/j.jse.2009.10.021>
15. Veehof MM, Slegers EJ, van Veldhoven NH, Schuurman AH, van Meeteren NL. Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). *J Hand Ther* 2002;15:347-54.
16. Wright BD, Linacre JM. Reasonable mean-square fit values. In: *Rasch Measurement Transactions* 1994;8. p. 370-1.
17. Wright BD, Masters G. Number of person or item strata. In: *Rasch Measurement Transactions*; 2002. p. 888. Available at: <http://www.rasch.org/rmt/rmt163f.htm>. Accessed April 04, 2019.
18. Wylie JD, Beckmann JT, Granger E, Tashjian RZ. Functional outcomes assessment in shoulder surgery. *World J Orthop* 2014;5:623-33. <https://doi.org/10.5312/wjo.v5.i5.623>