

Examination of the Applicability of the Disabilities of the Arm, Shoulder and Hand (DASH) Questionnaire to Patients with Hand Injuries and Diseases Using Rasch Analysis

Kathrin Braitmayer¹ · Caroline Dereskewitz² · Cornelia Oberhauser¹ · Klaus-Dieter Rudolf² · Michaela Coenen¹ 

Published online: 22 December 2016
© Springer International Publishing Switzerland 2016

Abstract

Background The Disabilities of the Arm, Shoulder and Hand (DASH) Questionnaire is the most commonly applied patient-reported outcome measure used to assess disability and functioning in clinical research and practice for patients with injuries and diseases of the upper extremities. The objective of this study was to assess whether the DASH is a valid and reliable questionnaire to measure disability and functioning in patients with hand injuries and diseases using Rasch analysis.

Methods We performed a psychometric study using data derived from two multicentre studies carried out to develop the International Classification of Functioning, Disability and Health (ICF) Core Sets for Hand Conditions. We analysed the data of 417 patients recruited in a clinical setting and suffering from hand injuries (e.g. finger fracture, flexor tendon injury) or diseases (e.g. carpal tunnel syndrome, rhizarthrosis). We examined whether the DASH fulfilled the assumptions for its use as a measure by applying a partial credit model and testing for differential item functioning for sex and age.

Results Bifactor analysis revealed problems with the underlying latent trait of functioning and disability. Rasch analysis raised further issues, including disordered

thresholds for eight items and misfit in nine items. One item showed Differential Item Functioning for sex.

Conclusion The study reveals that some DASH items do not fit the underlying trait that the DASH aims to measure. Further studies using Rasch analysis are needed to compare our findings with results of studies involving other target groups (e.g. patients with injuries of the upper arm and shoulder).

Key Points for Decision Makers

Our study cannot ultimately confirm an underlying trait of functioning and disability when applying the Disabilities of the Arm, Shoulder and Hand (DASH) Questionnaire to a sample of patients with hand injuries and diseases.

Especially patients with severe problems in functioning have difficulties differentiating between the response options of the DASH.

Further studies using Rasch analysis and including other (clinical) target groups with hand, arm and shoulder injuries or diseases are needed.

✉ Michaela Coenen
michaela.coenen@med.lmu.de

¹ Chair for Public Health and Health Services Research, Department of Medical Informatics, Biometry and Epidemiology-IBE, Research Unit for Biopsychosocial Health, Ludwig-Maximilians-Universität (LMU) Munich, Marchioninstr. 17, 81377 Munich, Germany

² Department of Hand Surgery, Plastic- and Microsurgery, BG Hospital Hamburg, Bergedorfer Straße 10, 21033 Hamburg, Germany

1 Introduction

Patient-reported outcome measures (PROMs) are widely accepted tools to measure health outcomes within clinical research and practice. Data and results retrieved from PROMs highly benefit the assessment of patients'

problems, their care, further assignment of interventions, and evaluation of research [1]. Thus, PROMs used for these purposes need to be robust, meaningful and interpretable, and in case of disease-specific PROMs, targeted to the needs and problems of patients affected by the respective health condition.

Several PROMs have been developed for the upper extremities and the hand and used for clinical and research purposes [2], such as the Michigan Hand Outcomes Questionnaire [3], the Western Ontario Rotator Cuff Index [4] and the Boston Carpal Tunnel Questionnaire [5]. The advantages of PROMs focusing on a defined anatomical region such as the hand or wrist (e.g. Michigan Hand Outcomes Questionnaire) instead of disease-specific PROMs (e.g. Boston Carpal Tunnel Questionnaire) have been revealed in terms of clinical practicability and applicability [6].

It is also a well-known fact that persons affected by injuries and diseases of the hand or the upper extremities have specific needs and problems in functioning and everyday life, such as problems in fine hand use, self-care, domestic life and social participation [7–9]. However, PROMs developed for upper extremities do not capture all functioning aspects—as defined by the World Health Organization's (WHO's) International Classification of Functioning, Disability and Health (ICF) [10]—that are relevant from the patient's perspective; this is pointed out by Coenen et al. [11].

The Disabilities of the Arm, Shoulder and Hand (DASH) Questionnaire [12] is the most commonly used PROM in clinical research and practice for patients with injuries and diseases of the upper extremities [13], and is available in more than 30 languages. It is a 30-item PROM aiming to measure functioning and disability in people with all kinds of injuries and diseases of the upper extremities. The DASH was developed by experts who selected items from a large pool of existing scales [12, 14]. The reliability and validity of the DASH have been thoroughly investigated by classical test theory approaches in various settings [14–18]. Internal reliability (Cronbach's alpha) has been reported as excellent in various subpopulations (0.92–0.98) [16]. The German version of the DASH [19] was validated using data from a sample of patients with shoulder pain [17] and showed high internal reliability (Cronbach's alpha) (0.96). Analyses addressing dimensionality and internal construct validity have rarely been conducted, and studies using factor analysis have shown varying results [20–23]. Unidimensionality was a problem in these four studies, and factor analysis supported a two-factor [23] or three-factor [20–22] model.

The strengths and weaknesses of PROMs can also be analysed by using Rasch analysis. Up to now this method has not been extensively used to analyse the DASH, apart

from the development of a shortened version [24] on the one hand, and studies focusing on patients with specific health conditions, namely multiple sclerosis [25] and musculoskeletal disorders of the upper extremities (e.g. shoulder arthritis) [21], on the other. The DASH was developed for persons with any kind of injury and disease of the upper extremities. However, it is also used for persons with only a specific anatomical region affected, such as for patients with hand injuries and diseases. This versatility generally offers a broader range of clinical and research applications, but might also bear the risk of the outcome being too unspecific for people with hand injuries and diseases.

The objective of this study was to examine whether the DASH is a valid and reliable questionnaire to assess functioning and disability of patients with hand injuries and diseases using Rasch analysis.

2 Methods

2.1 Design

This study is designed as a psychometric study using data from two multicentre cross-sectional studies that were carried out to develop and validate the ICF Core Sets for Hand Conditions [26, 27]. Both studies were conducted in a clinical setting in four German hospitals run by the German Statutory Accident Insurance. In these studies study participants with a broad range of injuries and diseases of the hand filled in the DASH amongst other RPOMs. Socio-demographic and disease-related data were collected by health professionals and study nurses, respectively. Sampling strategies, inclusion and exclusion criteria, as well as the data collection of these studies have been described elsewhere in detail [26–28]. Both studies were performed based on the principles of the Declaration of Helsinki and were approved by the Ethics Committee of the Medical Association, Hamburg, and the Ethics Committee of the Ludwig-Maximilians-Universität (LMU), Munich.

For this psychometric study we used bifactor and Rasch analysis to examine the reliability and validity of the DASH in persons with all kinds of hand injuries and diseases.

2.2 Data Used for Analyses

We used socio-demographic and disease-related data, as well as data derived from the German version of the DASH assessed in the two former empirical studies. Each of the 30 DASH items scores from 1 (e.g. 'no difficulty') to 5 (e.g. 'unable'). The 'DASH disability/symptom score' is

calculated by first averaging the given answers, then subtracting one and multiplying it by 25. The score ranges from 0 to 100, where a higher score indicates greater disability [14].

For our analyses, we solely considered data of patients who (a) had a valid DASH score, i.e. not more than three missing answers and (b) suffered from an injury (e.g. wrist fracture) or disease of the hand (e.g. carpal tunnel syndrome). Therefore, we excluded 29 patients with invalid DASH scores and 23 patients because of their diagnoses from the two samples ($N = 469$), resulting in a final sample of 417 patients used for further analysis.

2.3 Data Analyses

2.3.1 Descriptive Analysis

Descriptive analysis was carried out to describe the sample in terms of socio-demographic and disease-related characteristics. Results of the descriptive analysis are displayed in a table showing frequencies and percentages.

2.3.2 Internal Consistency Reliability

The internal consistency reliability of the DASH items was assessed based on different measures. For each item, the inter-item correlation and the item-total correlation (defined as the correlation between the item and the total DASH score) were calculated [29–32]. In addition, Cronbach's alpha and McDonald's omega total were computed.

2.3.3 Metric Analysis

We carried out metric analysis according to the conceptualization of the item response theory, starting with the testing of model assumptions and ending with Rasch analysis; additionally, we tested for Differential Item Functioning (DIF).

(1) Testing data for Rasch model assumptions: We tested for the model assumptions: unidimensionality, monotonicity and local independency. Unidimensionality was examined by means of a bifactor analysis [33, 34]. The assumption made by bifactor analysis is the existence of one general factor and multiple independent group factors. High loadings on the general factor for all items point to an underlying unidimensional latent trait. The number of factors considered in the bifactor analysis were determined based on permuted parallel analysis [35]. According to this method, the number of factors is defined as the number of eigenvalues resulting from the observed data exceeding the 95% quantile of the eigenvalues resulting from several permutations of the observed data. In addition, a single-factor model was estimated to compare the factor loadings

with the general factor from the bifactor analysis. If the loadings are similar, the general factor in the bifactor model represents the main dimension of the construct very well. Monotonicity was examined for each item by reviewing graphs of the item distributions conditional on average 'rest scores'. These scores were calculated for each person as the total raw score of all remaining non-missing items divided by their number. Monotonicity can be assumed if there is a consistent trend that persons with higher rest scores are more likely to have more problems in the given item. Local independency was examined based on residual correlations among items resulting from a single-factor analysis [36]. A residual correlation higher than $r = 0.3$ suggests that a response of one question influences the response to another question.

(2) Computing and fitting of the Rasch model: Rasch analysis generally explores to what extent the observed data, i.e. the actual given answers to the items, fits the Rasch model [37]. Within Rasch analysis, the model assumes an underlying latent trait; in this study, it is the trait of functioning and disability. On this trait, both items and persons are being located. The Partial Credit Model (PCM) was chosen to perform Rasch analysis in our study because of the set of ordinal, polytomous DASH items.

Within the PCM, the item thresholds for each item are calculated, indicating the location on the latent trait where the item best discriminates between persons [37–40]. Likewise, each person is assigned a value on the latent trait (called person's ability). The higher a person's ability is in relation to the threshold of an item, the higher the probability of a problem on that item. When a person's location on the latent trait is equal to the threshold of the item, there is by definition a 0.5 probability of a response below or above this threshold. In case of unordered thresholds, the response options have to be collapsed until they are in a correct order. If collapsing of response options was necessary for specific items in our analysis, the collapsing strategy was decided for each item separately. The person separation index was calculated for the final model.

Chi square-based item infit and outfit mean square statistics were computed. The outfit statistic includes differences between the observed and expected responses for all items. The infit statistic is weighted and emphasizes those items near the person's ability [37]. Values close to 1 indicate good item fit, while values larger than 1 demonstrate underfit (i.e. the observed data varies considerably more than can be explained by the model, constituting a violation of the model). Values smaller than 1 indicate overfit (i.e. the data varies noticeably less than expected based on the model, which is commonly considered acceptable) [41, 42]. According to the available literature, a range between 0.7 and 1.3 is generally acceptable [41].

(3) Testing for DIF: If an item shows DIF, it is a potential cause for bias in person measurement. This is the case where subgroups of a sample (e.g. older and younger patients) respond in a different way to a specific item despite equal levels of the underlying characteristic being measured. We tested DIF for sex and age dichotomized in above or below 45 years (thus splitting the sample in two almost equally sized groups at the median). Due to the sample size, change in McFadden's pseudo R^2 (>0.02) was chosen as the criterion for flagging.

Data analyses were performed using R software [43] and R package eRm [44] (Rasch analysis) as well as R package lordif [45] (testing for DIF).

3 Results

3.1 Sample

The age of the 417 patients included in the data analyses ranged from 18 to 82 years, with a mean age of 45.0 [standard deviation (SD) 13.7] years. Table 1 provides an overview of socio-demographic and disease-related characteristics of the sample. Detailed information on the patients' diagnoses is shown in Table 2. The DASH score ranged from 0 to 97.5, with a mean score of 41.4 (SD 21.3).

3.2 Internal Consistency Reliability

The DASH items showed high internal consistency reliability for our sample. Inter-item correlation was high, but

Table 1 Socio-demographic and disease-related characteristics of patients ($N = 417$)

Characteristics	<i>N</i>	%
Sex		
Female	133	31.9
Male	284	68.1
Setting		
Inpatient	294	70.5
Outpatient	123	29.5
Handedness ($n = 416$; $md\ n = 1$)		
Right-handed	380	91.3
Left-handed	36	8.7
Type of hand condition		
Injury	330	79.1
Disease	87	20.9
Affected hand = dominant hand ($n = 416$; $md\ n = 1$)		
No	177	42.5
Yes	197	47.4
Both hands affected	42	10.1

md missing data

showed great variability (mean 0.50, range 0.12–0.89). Item-total correlation was higher than the inter-item correlation, with less variation (mean 0.72, range 0.50–0.84). Cronbach's alpha was 0.97, and McDonald's omega total was 0.98. Therefore, all values indicated high internal consistency reliability in our sample.

3.3 Rasch Analysis

3.3.1 Rasch Model Assumptions

Unidimensionality: We derived three factors (displayed as blue, red and green curves in Fig. 1) from the permuted parallel analysis. Figure 1 shows high loadings for all items on the general factor (black curve). In general, the loadings were very close to the loadings from the single-factor model (dotted black curve). However, it must be noted that the items 24–28 concerning sensations (i.e. tingling, weakness, stiffness) and pain [i.e. pain (in general), pain when performing an activity] had higher loadings on the factor displayed as the red curve than on the general factor (black curve). So we assumed that these items did not fit the latent trait of functioning and disability perfectly in our sample. We still decided to compute the PCM including these items, but compared it to a model without the items referring to sensation and pain. Furthermore, the DASH items 2 ('writing') and 3 ('turning a key') (blue curve) as well as 10 ('carrying a shopping bag') and 11 ('carrying a heavy object') (green curve) each formed a factor, but with factor loadings below those of the general factor.

Monotonicity: Mosaic plots were evaluated to detect monotonicity. A general trend of a person with higher rest scores showing greater problems could be seen in all items. Therefore, monotonicity could be concluded.

Local independency: We calculated residual correlations among all items based on a single-factor analysis to check for possible local dependency. Overall, the results showed very low residual correlations, with three exceptions in accordance to the results from the bifactor analysis. A high residual correlation was found for items 2 ('writing') and 3 ('turning a key') ($r = 0.31$) as well as for items 10 and 11 ($r = 0.36$), both referring to carrying objects.

3.3.2 Computing of Partial Credit Models

As stated before, we decided to compute and compare two different PCMs: one with and one without the questions addressing pain and sensations (items 24–28). Since they were only slight differences, we decided to solely use a model including all items to perform further analyses. A closer inspection of the threshold patterns supported the use of the PCM. Before fitting the model, we found eight items with unordered thresholds. We collapsed the

Table 2 Diagnoses of patients ($N = 417$); multiple answers per patient possible

Diagnoses (ICD-10 code)	<i>N</i>	%
Fracture at wrist and hand level (S62)	81	19.4
Fracture of forearm (S52)	68	16.3
Injury of muscle and tendon at wrist and hand level (S66)	52	12.5
Dislocation, sprain and strain of joints and ligaments at wrist and hand level (S63)	47	11.3
Cellulitis of parts of limb (L03)	30	7.2
Open wound of wrist and hand (S61)	30	7.2
Burn and corrosion of wrist and hand (T23)	30	7.2
Algoneurodystrophy (M89)	22	5.3
Arthrosis (M19)	19	4.6
Mononeuropathies of upper limb (G56)	13	3.1
Synovitis and tenosynovitis (M65)	13	3.1
Injury of muscle and tendon at forearm level (S56)	11	2.6
Sequelae of injuries of upper limb (T92)	11	2.6
Superficial injury of wrist and hand (S60)	10	2.4

ICD-10 International Statistical Classification of Diseases and Related Health Problems, version 10

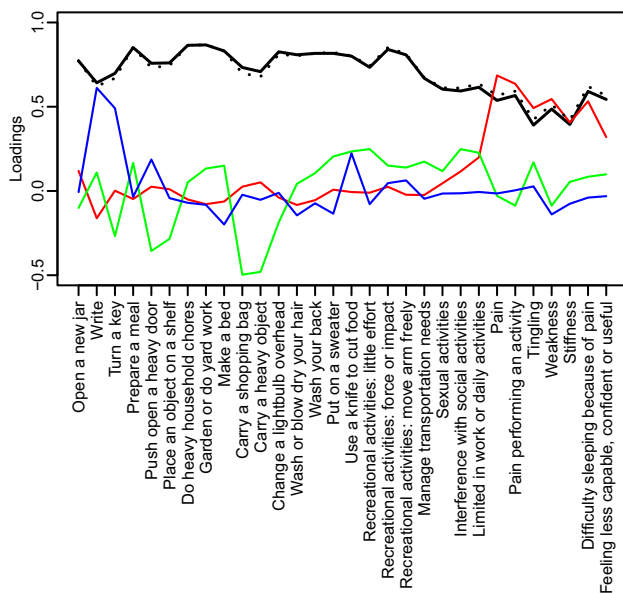


Fig. 1 Loadings retrieved from bifactor analysis. The solid black curve constitutes the factor loadings for the general factor, while the coloured curves (blue, red and green) constitute the factor loadings for the three group factors from the bifactor analysis. The dotted black curve contains the factor loadings from a single-factor model

response options in these until all thresholds were in the correct order (see Fig. 2). The final PCM was calculated and tested using the collapsed data. The person separation index for the model with the collapsed items was 0.96, yielding high reliability. Nine items showed misfit to the fitted PCM model. The infit and outfit mean square statistics for this model revealed an overfit for six items and an underfit for three items (see Table 3).

Figure 3 displays the person-item map. In the upper part, the grey histogram reflects the distribution of the

Items	Item labels	Collapsing strategies
2	Write	1 2 3 4 5
3	Turn a key	1 2 3 4 5
11	Carry a heavy object	1 2 3 4 5
12	Change a light bulb over head	1 2 3 4 5
19	Recreational activities: move arm freely	1 2 3 4 5
21	Sexual activities	1 2 3 4 5
28	Stiffness	1 2 3 4 5
30	Feel less capable, confident or useful	1 2 3 4 5

Fig. 2 Collapsing strategies of items with unordered thresholds. DASH items with unordered thresholds of the response options ranging from 1 to 5 are displayed along with the respective collapsing strategies. Collapsed response options are marked by circles

persons' abilities on the latent trait of functioning and disability. Persons' abilities range from -4.41 to 4.24 (mean -0.09 , median 0.01) (see x -axis values). In the lower part, the item thresholds (black circles) are visualized. The item thresholds cover most of the range (from -1.94 to 3.94 , mean 0.41 , median 0.25), but leave gaps, especially at the lower end of the continuum. However, only a very small part of the sample (25 persons, i.e. 6.0%) has persons' abilities below the lowest threshold. To facilitate judgment on targeting between persons' abilities and item thresholds, the item thresholds are additionally visualized by little black lines directly beneath the histogram of the persons' abilities in the upper part of Fig. 3. Overall, the items well target the persons.

Table 3 Polychoric correlations of the DASH items with the rest score and item fit statistics, in order of DASH items

Item	Item labels	Correlation with rest score	Infit MSQ	Outfit MSQ
1	Open a new jar	0.77	0.86	0.87
2	Write	0.62	1.07	1.11
3	Turn a key	0.71	0.91	0.83
4	Prepare a meal	0.83	0.66 ^a	0.64 ^a
5	Push open a heavy door	0.75	0.93	0.89
6	Place an object on a shelf	0.76	0.93	0.91
7	Do heavy household chores	0.84	0.65 ^a	0.64 ^a
8	Garden or do yard work	0.83	0.66 ^a	0.65 ^a
9	Make a bed	0.80	0.72	0.69 ^a
10	Carry a shopping bag	0.73	1.05	1.05
11	Carry a heavy object	0.71	0.93	0.89
12	Change a lightbulb overhead	0.81	0.74	0.69 ^a
13	Wash or blow dry your hair	0.79	0.79	0.73
14	Wash your back	0.80	0.80	0.81
15	Put on a sweater	0.82	0.69 ^a	0.64 ^a
16	Use a knife to cut food	0.79	0.80	0.77
17	Recreational activities: little effort	0.73	0.96	0.94
18	Recreational activities: force or impact	0.83	0.76	0.77
19	Recreational activities: move arm freely	0.79	0.83	0.82
20	Manage transportation needs	0.68	1.17	1.46
21	Sexual activities	0.63	1.07	1.25
22	Interference with social activities	0.64	1.14	1.40
23	Limited in work or daily activities	0.66	1.00	0.98
24	Pain	0.64	1.11	1.14
25	Pain performing an activity	0.66	1.07	1.06
26	Tingling	0.50	1.43 ^b	1.57 ^b
27	Weakness	0.59	1.30	1.37 ^b
28	Stiffness	0.51	1.34 ^b	1.36 ^b
29	Difficulty sleeping because of pain	0.68	1.04	1.01
30	Feeling less capable, confident or useful	0.62	1.06	1.11

DASH Disabilities of the Arm, Shoulder and Hand Questionnaire, MSQ mean square

^a Items showing overfit

^b Items showing underfit

3.3.3 Testing for Differential Item Functioning

No DIF was found for age. Item 11 ('carrying a heavy object') showed DIF for gender.

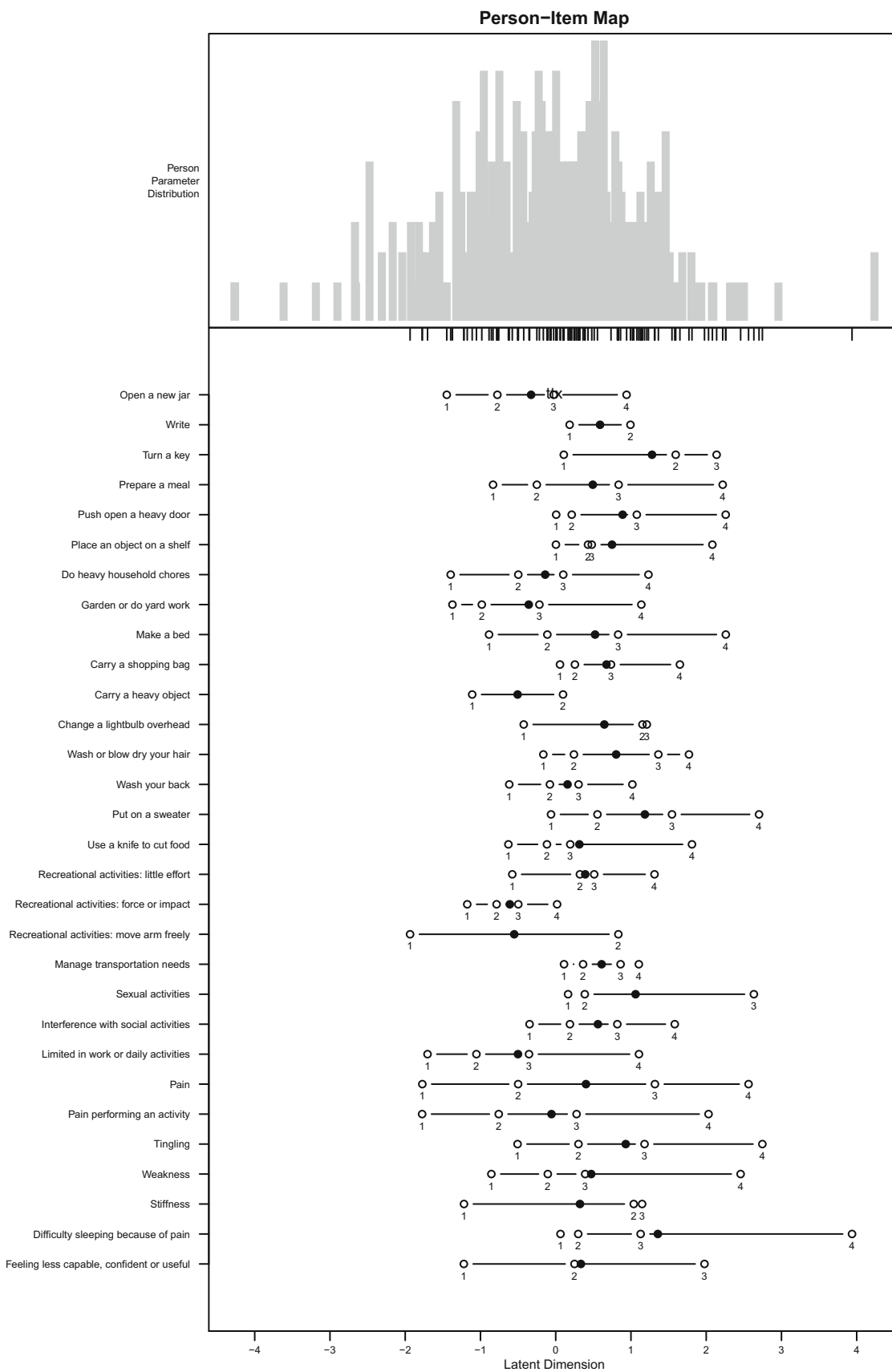
4 Discussion

With this psychometric study we examined whether the DASH is a valid and reliable questionnaire to measure functioning and disability in patients with hand injuries and diseases using Rasch analysis. Previous studies using traditional psychometric methods in different populations and various settings showed excellent results on reliability and validity. Our study cannot ultimately confirm an underlying

Fig. 3 Person-item map. The histogram in the *upper part* displays the distribution of the persons' abilities. In the *lower part*, the *circles* illustrate the thresholds for each response category of the respective items, while (as a summary measure) the *black dots* indicate the mean of the thresholds for the respective item. The item thresholds are in addition visualized by little *black lines* directly beneath the histogram of persons' abilities

latent trait when applying the DASH to a sample of patients with hand injuries and diseases. Some items also showed disordered response options and item misfit.

The DASH was developed and is used for persons suffering from any kind of injury and disease of the upper extremities. As it aims to measure the extent of disability of a person, all items should be located on a latent trait or continuum of disability. From a methodological point of



view, there is no standard approach to prove this continuum and more specifically to examine whether a PROM is unidimensional or not. In our analysis we decided to examine unidimensionality using bifactor analysis. Our bifactor analysis showed that items referring to pain and sensations did not fit the underlying dimension of disability. Although pain and sensations are clinically relevant and important aspects of disability for patients, the inclusion of these items into the overall DASH score seems to be problematic. The residual correlations and the bifactor analysis both revealed problematic results for items 2 ('write') and 3 ('turn a key'). According to Reise this can be explained by the content similarity of these two items [34].

We found that nine out of 30 items misfit the model. Particularly, activities such as preparing a meal (item 4), doing heavy household chores (item 7) or making a bed (item 8) did not fit the model. This could have been caused by our sample including mostly men who probably do not perform these activities on a regular basis. Even though it showed no misfit, writing (item 2) was also a problematic item. After discussions within our research team, we concluded that writing as an activity may not be of relevance in the digital age or for patients with the working background of a craftsperson, of which our sample mostly consisted. Furthermore, we found DIF for item 11 ('carry a heavy object'). Yet, one must consider that even without the existence of a hand injury or disease, men and women may not respond to this question similarly.

The items referring to recreational activities partly showed unordered thresholds before fitting the final PCM. This might be explained by the fact that activities such as knitting, golfing or playing Frisbee do not reflect hobbies carried out by the majority of the population. In addition, it seemed to be difficult for our study participants (most of them treated as inpatients during data collection) to rate activities referring to recreation and leisure, as patients in clinical settings hardly perform them. This should be considered when applying the DASH as an outcome measure in clinical trials and in clinical settings. The issues reported beforehand might hint at the possibly limited conceptual base of the DASH. In the development process of the DASH, no psychometric techniques were applied for item reduction [12]. Items were selected and added on the basis of experts' opinions. Therefore, items did not necessarily measure the same construct or operationalize important aspects as intended.

Besides the fit of items, it is important to look at their discrimination ability in further detail. We found disordered thresholds for eight items. This could have several reasons. The most intuitive one is that the persons filling in the DASH had difficulties discriminating between the five response options and their labels. Detailed investigations of

the item characteristic curves showed that it seemed to be difficult for patients to differentiate between 'moderate' and 'severe difficulty', as well as 'severe difficulty' and 'unable'. Cano and colleagues also reported similar findings in their sample of patients with multiple sclerosis [25]. One option to overcome this problem is to reconsider the labelling or even the number of the response options; especially since the wording of the response options is only identical for the first 21 questions (i.e. 'no', 'mild', 'moderate' or 'severe difficulty', 'unable') and then changes four times to, for example, 'not at all', 'slightly', 'moderately', 'quite a bit' and 'extremely' in item 22.

With regard to the person parameters, our findings suggest that the DASH is not perfectly targeted for people who suffer from injuries and diseases exclusively located on the hand. Both the mean location for persons and items on a well-targeted PROM (i.e. not too hard, nor too easy) should be around the value of zero [40]. In our study, we detected a slight shift to the right for the item thresholds. Accordingly, the persons' abilities of a small part of the sample could not be differentiated optimally on the basis of the available item thresholds at the lower end of the continuum. However, this should be interpreted with caution as we used data from a clinical sample including a number of patients with severe impairments [7, 9, 11].

This study has some limitations. Firstly, the data used for our analyses was collected in German hospitals run by the German Statutory Accident Insurance. Thus, many patients of our sample had injuries following accidents at work (e.g. cutting damage with transection of a nerve), which might not reflect the whole range of hand injuries in general and might have biased the results. Secondly, due to the typical clientele of those hospitals, our sample showed an unbalanced gender proportion, which might have influenced the DIF testing. Thirdly, it remains unclear at which treatment stage the patients were included in the two former studies and more specifically when they filled out the DASH. This convenience, clinical sample might cause some limitations with regard to the generalizability of our results to other patient groups or settings. Fourthly, using the median age of 45 as the cut-off to test for DIF was an arbitrary decision. Finally, it is possible that some linguistic and cultural characteristics of the German version of the DASH could have influenced some of our results.

5 Conclusions

Our study reveals that some DASH items do not fit the underlying trait of disability that the DASH aims to assess. Unordered thresholds of items shed light on patients' problems in differentiating between response options when filling out the DASH, especially in those who are severely

affected. It would be worthwhile to perform further studies using Rasch analysis with other target groups (e.g. patients with injuries of the upper arm and shoulder) to compare their findings with our results.

Acknowledgements We would like to thank all persons involved in the data collection at (1) the Department of Hand Surgery, Plastic and Microsurgery at the BG Klinikum Hamburg, (2) the Department of Plastic and Hand Surgery, Burns Unit at the BG Klinikum Bergmannstrost Halle, (3) the Department of Hand Surgery, Plastic and Microsurgery at the Unfallkrankenhaus Berlin and (4) the Department of Hand, Plastic, Reconstructive and Burns Surgery at the BG Klinik Tübingen.

Author contributions Michaela Coenen initiated and supervised the study, contributed to the design of the study and revised the draft version of the manuscript. Kathrin Braitmayer performed the data analysis, interpreted the results and drafted the manuscript. Cornelia Oberhauser supervised the data analysis and contributed to the writing and revising of the manuscript. Caroline Dereskewitz and Klaus-Dieter Rudolf contributed to the design of the study, managed and supervised the data collection of the two former studies and contributed to the writing of the manuscript.

Compliance with Ethical Standards

Funding The two former studies that provided the data used in this study were funded by the German Social Accident Insurance, DGUV. The authors of this publication are responsible for its content.

Disclosure of potential conflicts of interest K Braitmayer, C. Dereskewitz, C. Oberhauser, K.-D. Rudolf and M. Coenen declare that they have no competing interests.

Statement on human rights All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Statement on the welfare of animals This article does not include any animal studies performed by any of the authors.

Informed consent For this psychometric study, we used data from two former studies. Both studies were approved by the Ethics Committee of the Medical Association, Hamburg, and the Ethics Committee of the Ludwig-Maximilians-Universität (LMU), Munich. Informed consent was obtained from all individual participants included in the study.

References

- Hobart JC, Cano SJ, Zajick JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol.* 2007;6(12):1094–105.
- Smith MV, Calfee RP, Baumgarten KM, Brophy RH, Wright RW. Upper extremity-specific measures of disability and outcomes in orthopaedic surgery. *J Bone Joint Surg Am.* 2012;94(3):277–85.
- Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am.* 1998;23(4):575–87.
- Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: the Western Ontario Rotator Cuff Index. *Clin J Sport Med.* 2003;13(2):84–92.
- Levine DW, Simmons BP, Koris MJ, Daltroy LH, Hohl GG, Fossel AH, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am.* 1993;75(11):1585–92.
- Davis AM, Beaton DE, Hudak P, Amadio P, Bombardier C, Cole D, et al. Measuring disability of the upper extremity: a rationale supporting the use of a regional outcome measure. *J Hand Ther.* 1999;12(4):269–74.
- Bialocerkowski AE. Difficulties associated with wrist disorders—a qualitative study. *Clin Rehabil.* 2002;16(4):429–40.
- Wong JY. Time off work in hand injury patients. *J Hand Surg Am.* 2008;33(5):718–25.
- Hill S, Dziedzic KS, Ong BN. The functional and psychological impact of hand osteoarthritis. *Chronic Illn.* 2010;6(2):101–10.
- World Health Organization. *International Classification Of Functioning, Disability and Health.* Geneva: World Health Organization; 2001.
- Coenen M, Kus S, Rudolf KD, Muller G, Berno S, Dereskewitz C, et al. Do patient-reported outcome measures capture functioning aspects and environmental factors important to individuals with injuries or disorders of the hand? *J Hand Ther.* 2013;26(4):332–42 (quiz 42).
- 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(6):602–8.
- Kus S, van de Ven-Stevens LA, Coenen M, Berno S, Kollerits B, Cieza A. What is our knowledge of functioning and disability in hand conditions based on? *Arch Phys Med Rehabil.* 2011;92(8):1326–32.
- Beaton DE, Davis AM, Hudak PL, McConnell S. The DASH (Disabilities of the Arm, Shoulder and Hand) outcome measure: what do we know about it now? *Br J Hand Ther.* 2001;6(4):109–18.
- Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther.* 2001;14(2):128–46.
- Gummesson C, Atroshi I, Ekdahl C. The Disabilities of the Arm, Shoulder and Hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. *BMC Musculoskelet Disord.* 2003;4:11–6.
- Offenbaecher M, Ewert T, Sangha O, Stucki G. Validation of a German version of the Disabilities of Arm, Shoulder, and Hand Questionnaire (DASH-G). *J Rheumatol.* 2002;29(2):401–2.
- SooHoo NF, McDonald AP, Seiler JG 3rd, McGillivray GR. Evaluation of the construct validity of the DASH questionnaire by correlation to the SF-36. *J Hand Surg Am.* 2002;27(3):537–41.
- Germann G, Harth A, Wind G, Demir E. Standardisation and validation of the German version 2.0 of the Disability of Arm, Shoulder, Hand (DASH) Questionnaire. *Unfallchirurg.* 2003;106(1):13–9.
- Lehman LA, Woodbury M, Velozo CA. Examination of the factor structure of the Disabilities of the Arm, Shoulder, and Hand Questionnaire. *Am J Occup Ther.* 2011;65(2):169–78.
- Franchignoni F, Giordano A, Sartorio F, Vercelli S, Pascariello B, Ferriero G. 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(9):1370–7.

22. Cheng HM, Sampaio RF, Mancini MC, Fonseca ST, Cotta RM. Disabilities of the Arm, Shoulder and Hand (DASH): factor analysis of the version adapted to Portuguese/Brazil. *Disabil Rehabil.* 2008;30(25):1901–9.
23. Mousavi SJ, Parnianpour M, Abedi M, Askary-Ashtiani A, Karimi A, Khorsandi A, et al. Cultural adaptation and validation of the Persian version of the Disabilities of the Arm, Shoulder and Hand (DASH) outcome measure. *Clin Rehabil.* 2008;22(8):749–57.
24. Beaton DE, Wright JG, Katz JN, Upper Extremity Collaborative G. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am.* 2005;87(5):1038–46.
25. Cano SJ, Barrett LE, Zajicek JP, Hobart JC. Beyond the reach of traditional analyses: using Rasch to evaluate the DASH in people with multiple sclerosis. *Mult Scler.* 2011;17(2):214–22.
26. Rudolf KD, Kus S, Chung KC, Johnston M, LeBlanc M, Cieza A. Development of the international classification of functioning, disability and health core sets for hand conditions—results of the World Health Organization International Consensus process. *Disabil Rehabil.* 2012;34(8):681–93.
27. Kus S, Oberhauser C, Cieza A. Validation of the brief international classification of functioning, disability, and health (ICF) core set for hand conditions. *J Hand Ther.* 2012;25(3):274–86 (quiz 87).
28. Kus S, Dereskewitz C, Wickert M, Schwab M, Eisenschenk A, Steen M, et al. Validation of the comprehensive international classification of functioning, disability and health (ICF) core set for hand conditions. *Hand Therapy.* 2011;16(3):58–66.
29. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16(3):297–334.
30. Revelle W, Zinbarg R. Coefficients alpha, beta, omega, and the glb: comments on Sijtsma. *Psychometrika.* 2009;74(1):145–54.
31. Roscino A, Pollice A. A generalization of the polychoric correlation coefficient. In: Zani S, Cerioli A, Riani M, Vichi M, editors. *Data analysis, classification and the forward search.* Berlin, Heidelberg: Springer; 2006. p. 135–42.
32. Trochim W. *The research methods knowledge base*, 2nd edition. Internet WWW page (version current as of October 20, 2006) 2016 [updated October 20, 2006; cited 2016 March 22, 2016]. <http://www.socialresearchmethods.net/kb/>.
33. Jennrich RI, Bentler PM. Exploratory bifactor analysis. *Psychometrika.* 2011;76(4):537–49.
34. Reise SP. The rediscovery of bifactor measurement models. *Multivar Behav Res.* 2012;47(5):667–96.
35. Buja A, Eyuboglu N. Remarks on parallel analysis. *Multivar Behav Res.* 1992;27(4):509–40.
36. Reeve BB, Fayers P. Applying item response theory modeling for evaluating questionnaire item and scale properties. In: Fayers P, Hays RD, editors. *Assessing quality of life in clinical trials: methods of practice.* 2nd ed. Oxford: Oxford University Press; 2005. p. 55–73.
37. Elhan AH, Küçükdeveci AA, Tennant A. The Rasch measurement model. In: Franchignoni F, editor. *Research issues in physical and rehabilitation medicine.* Pavia: Maugeri Foundation Books; 2010. p. 89–102.
38. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *J Rehabil Med.* 2003;35(3):105–15.
39. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum.* 2007;57(8):1358–62.
40. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol.* 2007;46(Pt 1):1–18.
41. Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. *BMC Med Res Methodol.* 2008;8(1):33.
42. Bond TG, Fox CM. *Applying the Rasch model: fundamental measurement in the human sciences.* 2nd ed. Mahwah: Lawrence Erlbaum Associates, Inc.; 2007.
43. R Core Team. *R: a language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing; 2013.
44. Mair P, Hatzinger R, Maier MJ. *eRm: Extended Rasch Modeling.* 0.15-7. 2016. <http://erm.r-forge.rproject.org/>. Accessed 20 Dec 2016.
45. Choi SW, Gibbons LE, Crane PK. lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *J Stat Softw.* 2011;39(8):1–30.