The structural validity of the Finnish version of the Disabilities of the Arm, Shoulder and Hand: A Rasch model analysis

Hand Therapy

Ikonen et al.

J Ikonen[1,2] S Hulkkonen[3] J Ryhänen[3] A Häkkinen[4] J Karppinen[5,6] JP Repo[1]

[1]. Department of Surgery, Central Hospital of Central Finland, Jyväskylä, Finland
[2]. Department of Hand Surgery, Turku University Hospital and University of Turku, Turku, Finland
[3]. Department of Hand Surgery, Helsinki University Hospital and University of Helsinki, Helsinki, Finland
[4]. Faculty of Sports and Health Sciences, University of Jyväskylä, Jyväskylä, Finland
[5]. Medical Research Center Oulu, Oulu University Hospital, University of Oulu, Oulu, Finland
[6]. Finnish Institute of Occupational Health, Oulu, Finland

J Ikonen, Tehtaankatu 6 as 25, 20500 Turku, Finland. Email: jooiko@utu.fi[AQ1]

ABSTRACT

The construct validity of the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH) has previously been questioned. The purpose of this study was to evaluate the measurement properties of the Finnish version of the DASH for assessing disability in patients with hand complaints using Rasch Measurement Theory.

A cohort of 193 patients with typical hand and wrist complaints were recruited at a surgery outpatient clinic. The DASH scores were analysed using the Rasch model for differential item functioning, unidimensionality, fit statistics, item residual correlation, coverage/targeting and reliability.

In the original DASH questionnaire, the item response thresholds were disordered for 2 of 30 of the items. The item fit was poor for 9 of 30 of the items. Unidimensionality was not supported. There was substantial residual correlation between 87 pairs of items. Item reduction (chi square 95, degrees of freedom 50, p < 0.001) and constructing two testlets led to unidimensionality (chi square 0.64, degrees of freedom 4, p = 0.96). Person separation index was 0.95. The testlets had good fit with no differential item functioning towards age or gender.

Unidimensionality of the original Finnish version of the DASH was not supported, meaning the questionnaire seems to gauge traits other than disability alone. Hence, the clinician must be careful when trying to measure change in patients' scores. Item reduction or the creation of testlets did not lead to good alternatives for the original Finnish DASH. Differential item functioning showed that the original Finnish scale exhibits minor response bias by age in ~~few~~one i~~s~~. The original Finnish DASH covers different levels of ability well among typical hand surgery patients.

Keywords: DASH , validation , Rasch , Finnish , disability , outcome

Introduction

Pain and disability are the predominant complaints faced by hand surgeons in the clinic. Patient-reported outcome measures (PROM) have been developed to help describe the disability experienced by patients. They can also be used to monitor changes in symptoms and function over time. Also, increasingly, clinical trials use PROMs as primary outcomes to assess the effectiveness of different treatment modalities.[1]

The Disabilities of the Arm, Shoulder and Hand (DASH) is a 30-item, self-administered questionnaire specific for the upper limb. It was developed as a joint initiative by the American Academy of Orthopedic Surgeons (AAOS), the Council of Musculoskeletal Specialty Societies (COMSS) and the Institute for Work and Health (Toronto, Canada).[2] Compared to questionnaires specific to a single joint or disorder, the DASH can be used for single and multiple disorders in the upper limb. Previous studies using classical test theory have placed the internal consistency of the DASH at 0.97–0.98.[3,4] Studies using item response theory models have questioned the structural validity of the DASH for patient groups with specific disorders.[5–7]

In 2008, a steering group translated and cross-culturally adapted the DASH into Finnish.[8] Scores on the Finnish version of the DASH have been shown to correlate well with the Michigan Hand Questionnaire and the Euro-Qol-5D-3L.[9] In adapting PROMs, it is important to ensure that the items are understood in the same way and that each item of the questionnaire represents a similar level of difficulty across different languages. Although widely used in clinical trials and in patient care alike, the construct validity of the Finnish version of the DASH has not yet been demonstrated in the literature.

Items in the DASH are scored on an ordinal scale. In a well-functioning PROM, the items behave at an interval-level scaling.[1] The Rasch analysis tests an outcome scale against a mathematical model.[10] It shows what should be expected in responses to items if interval scale measurement is to be achieved. The Rasch model should be used to validate a scale when the items in a questionnaire are intended to be summed together to provide a total score.[10,11] If the items do not behave at an interval-level scaling, they cannot accurately be summed together.

Studies that have used Rasch analysis on the DASH have shown several problems including multidimensionality of the data, disordered thresholds, differential item functioning (DIF) and high residual correlation between items.[12–14] Previously, to correct the problems of multidimensionality and local dependence, studies have removed items that do not fit the Rasch model[7,15] or split the DASH into subscales.[13,16,17] More recently, the creation of testlets has been proposed to solve these issues.[5]

These attributes are also of clinical importance. Analysing the structural validity of the DASH gives the surgeon or the therapist insight into the limitations of PROMs. Item response theory may for example give clues to whether patients are unable to answer some items just because the scale used does not lend itself well to the item (item thresholds) or whether some items measure differently according to age or sex (DIF).

In this study, the psychometric properties of the Finnish version of the DASH are tested using the Rasch Measurement Theory. The sample data are analysed for DIF, unidimensionality, fit statistics, item residual correlation, coverage/targeting and reliability.

Methods

The study protocol was approved by the Ethics Committee of the Northern Ostrobothnia Hospital District. The study adhered to the good ethical principles of the Declaration of Helsinki (World Medical Association 2013).[AQ2] The study invited 250 consecutive patients who had various hand and wrist complaints treated in the general orthopaedic outpatient clinic at Länsi-Pohja's Central Hospital, Kemi, Finland, to participate. Before arriving to the outpatient appointment, the patients completed the Finnish version of the Disabilities of the Arm, Shoulder and Hand (DASH). Adult patients who agreed to participate, whose first language was Finnish, had no cognitive dysfunction and could understand written Finnish were included in the present study. The patients were then debriefed and clinically examined by a surgeon. Exclusion criteria for the study were age of less than 18 years old and equal to or more than three unanswered items on the Finnish DASH. Patients with fractures of the hand and wrist were examined during their final follow-up appointment at the hospital. For all patients, both hands were examined. If both hands were affected, the worse one in the patient's opinion was included in the final analysis. Basic demographic and clinical information, such as handedness, occupation and other illnesses, was obtained from the patients.

Instruments

The DASH is a 30-item upper extremity-specific outcome measure. The items of the DASH can be divided into two distinct sections: physical activities (23 items) and symptoms (7 items). These items assess the overall state of the upper extremity during the preceding week. The DASH scores on a 5-point ordinal Likert scale (1 = no difficulty to 5 = unable). The sum of the answers is calculated and averaged, producing a score out of five. To create a scale from 0 to 100, this score is subtracted by one and multiplied by 25. Likert scaling assumes that the interval between each item response is equal. The DASH has two optional modules: work and sports/performing arts. These two optional modules were excluded from the analysis.

Statistical methods

Rasch measurement analysis[18] was used to measure construct validity, fit and reliability. Analyses included statistical and illustrative tests. The model and its psychometric criteria are described elsewhere in more detail.[9,19–23]

Log residuals (item–person interaction), chi-square (χ^2)-values (item–trait interaction) and item characteristic curves were analysed to evaluate item fit.[24,25] Fit residual indicates the proportion of the logit that exceeds the optimal fit value, and values between –2.5 and +2.5 are generally considered acceptable. Values not fitting into this range can be considered to indicate probability of inaccuracy, weakness in reliability or poorer quality of the testing. Non-significant p-values were hypothesized to be found after Bonferroni correction.

Residual correlations between each two items were examined to obtain information about the local dependency of items. A generally acknowledged threshold of equal or above 0.2 was used to recognize residual correlations. The authors hypothesized that the residual correlation would be below this threshold. High local dependency is also thought to artificially inflate the reliability of a test.[11]

The DASH is designed to assess a single trait, disability; thus, its construct should be unidimensional. According to general principles, the criterion for unidimensionality is less than 5% of significant t-tests, referring to the level of total variance.[26] Unidimensionality for this study was tested for item subsets with item residual loading of ±0.2. The authors hypothesized that the Finnish DASH would fulfil this criterion for unidimensionality measuring only one latent trait. If violation of the Rasch model for unidimensionality occurred, item reduction and testlet formation were conducted in separate analyses. Testlets can be created based on item residual correlations. Total non-error variance indicating unidimensional factor on which all items would load was set at a threshold of 85% as previously described by Prodinger et al.[5]

DIF was also examined. DIF occurs when subgroups, e.g. male/female, young/old, respond to a particular item differently even though they are at the same level of underlying ability.

The person separation index (PSI)[23] was measured to assess the level of reliability for the DASH. The resulting value is between 0 and 1, indicating lower or higher reliability, respectively. The authors hypothesized that the PSI-value would be at least 0.80.

Person and item locations were evaluated to determine whether the distribution of items matched with the coverage of the degree of disability measured in the study sample.[27] Scale targeting provided information into the suitability of the scale in this population of patients with hand disorders. The authors hypothesized that the Finnish DASH questionnaire provides good coverage and targeting for patients with hand complaints.

The SPSS 25.0 (IBM® SPSS® Statistics, USA), R-3.4.2 and Rumm2030 software packages were used for statistical analysis. Results are reported adhering to the STROBE and COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklists.[28]

Results

A total of 230 patients (participation rate, 92%) agreed to take part. Out of 230 patients, 193 (84%) had completed at least 27 items in the DASH and were included in the final analysis.

The most common diagnosis was carpal tunnel syndrome, affecting 42% of the study patients. Of the patients, 80 had no comorbidities and 113 had at least one chronic disease. Some patients had two concurring hand diagnosis and more than one chronic disease. The basic demographic features of the study patients are shown in Table 1.

Table 1. Sample demographics and patient diagnosis.

Variable	All, N = 193 (%)
Age, years, mean (SD, range)	54 (15, 19–91)
Handedness, n (%)	
Right	171 (88)
Left	16 (8)
Both	6 (4)
Analysed hand, n (%)	
Right	109 (56)
Left	84 (44)
Diagnosis of the analysed hand, n (%)	
Carpal tunnel syndrome	82 (42)
Ulnar nerve entrapment	3 (2)
Trigger finger	25 (13)
Ganglion cyst	17 (9)
Dupuytren's disease	17 (9)
CMC1 arthrosis	16 (8)
Distal radius fracture	20 (10)
Other fracture of the hand/wrist	20 (10)
Other	10 (5)
Comorbidities, n (%)	
None	80 (41)
Cardiovascular	48 (25)
Diabetes	25 (13)
Musculoskeletal	42 (22)
Neurological	12 (6)
Lung	7 (4)
Rheumatoid arthritis	8 (4)

Item fit and unidimensionality

The overall mean (SD) item fit residual was 0.02 (2.29). The overall person fit residual was –0.24 (1.53). Total item χ^2 was 169 (60 degrees of freedom; $p < 0.001$). Fit residuals were within ±2.5 for 21/30 of the items. Nine Items (2, 4, 8, 13, 21, 26, 28, 29 and 30) had a fit residual outside these parameters (Table 2). χ^2 values after Bonferroni correction were statistically significant ($p < 0.001$) for items 26 ("Tingling") and 29 ("Difficulty sleeping"). Residual correlation was above 0.2 for 87 pairs of items.[AQ3] Items 24 "Arm, shoulder or hand pain" and 25 "Arm, shoulder or hand pain when you performed any specific activity" had the highest residual correlation of 0.691. Items 20 ("Manage transportation needs") and 21 ("Sexual activities") showed disordered thresholds. Collapsing response categories 2 ("Mild difficulty"), 3 ("Moderate difficulty") and 4 ("Severe difficulty") into one category, "Difficult" led into ordered thresholds for these items.

Table 2. Item locations, fit residuals and differential item functioning of the original Finnish DASH.

Item	Location	Fit residual	Degrees of freedom	χ^2	Probability	Missing, n (%)	DIF by Age	DIF by gender
1. Open a tight or new jar	–0.86	–0.63	173.82	0.85	0.564	**3 (1.6)**	–	–

Item	Location	Fit residu-al	Degrees of free-dom	χ^2	Proba-bility	Missing, n (%)	DIF by Age	DIF by gender
2. Write	0.61	2.66	174.77	9.02	0.025	**2 (1.0)**	–	–
3. Turn a key	0.80	0.28	175.71	1.12	0.665	**1 (0.5)**	–	–
4. Prepare a meal	0.97	–2.63	174.77	6.97	0.004	**2 (1.0)**	–	**UDIF***
5. Push open a heavy door	0.78	–1.56	176.66	2.13	0.215	**0 (0)**	**UDIF****	–
6. Place an object on a shelf above your head	0.15	–0.10	172.88	1.30	0.590	**4 (2.1)**	–	–
7. Do heavy household chores	–0.56	–2.37	176.66	5.15	0.016	**1 (0.5)**	–	–
8. Garden or do yard work	–0.55	–2.52	175.71	4.07	0.030	**2 (1.0)**	–	–
9. Make a bed	0.85	–1.70	176.66	4.88	0.036	**1 (0.5)**	–	–
10. Carry a shopping bag	0.13	–0.14	175.71	0.12	0.941	**1 (0.5)**	–	–
11. Carry a heavy object (over 10 lbs)	–0.36	–1.91	176.66	1.38	0.501	**0 (0)**	–	**UDIF***
12. Change a lightbulb over-head	–0.58	–0.77	175.71	0.31	0.859	**1 (0.5)**	–	**UDIF****
13. Wash or blow dry your hair	0.73	–2.83	174.77	13.58	0.001	**2 (1.0)**	–	**UDIF***
14. Wash your back	–0.15	–1.02	175.71	2.29	0.320	**1 (0.5)**	–	–
15. Put on a pullover sweater	1.31	–1.51	176.66	5.42	0.067	**0 (0)**	–	–
16. Use a knife to cut food	0.63	–0.65	174.77	5.94	0.051	**2 (1.0)**	–	–
17. Recreational activities which require little effort	–0.18	0.20	173.82	0.65	0.724	**3 (1.6)**	–	–
18. Recreational activities with force or impact	–1.10	–1.72	171.93	3.88	0.144	**5 (2.6)**	–	**UDIF***
19. Recreational activities which move arm freely	–0.72	–1.43	170.05	1.20	0.369	**7 (3.6)**	–	–
20. Manage transportation needs	0.66	–0.55	175.71	0.54	0.764	**1 (0.5)**	–	–
21. Sexual activities	0.74	2.74	167.21	1.64	0.441	**10 (5.2)**	–	–
22. Interference with normal social activities	0.78	0.53	176.66	0.12	0.943	**0 (0)**	–	–
23. Limitation in work or other daily activities	–0.56	–1.56	173.82	3.88	0.144	**3 (1.6)**	–	–
24. Arm, shoulder or hand pain	–0.44	0.40	169.10	1.56	0.458	**8 (4.1)**	–	–
25. Pain performing specific activity	–0.97	0.13	171.93	0.02	0.991	**5 (2.6)**	–	–
26. Tingling	–0.30	5.19	172.88	63.12	0.000	**4 (2.1)**	–	–
27. Weakness	–0.61	1.77	173.82	3.56	0.169	**3 (1.6)**	–	–
28. Stiffness	–0.35	3.15	173.82	4.12	0.128	**3 (1.6)**	**UDIF***	–
29. Difficulty sleeping	–0.22	6.38	174.77	15.28	0.000	**2 (1.0)**	–	–
30. I feel less capable, confi-dent or useful	–0.62	–2.77	173.82	4.14	0.12	**3 (1.6)**	–	–

Note: Items in testlet 1 are shown in italics and items in testlet 2 in shown in bold.

DIF: differential item functioning; UDIF: uniform differential item functioning; NUD: non-uniform differential item functioning.

*p < 0.01; **p < 0.001.

Two subsets of six items were used for analysing unidimensionality (assessing a single latent trait, disability) for the Finnish DASH. These items exhibited item residual loading exceeding ±0.2. Unidimensionality was not supported, as there were 27.1% of significant t-tests (Table 2).

Item reduction

Separate analyses with item reduction and testlet formation were done in order to achieve unidimensionality. For item reduction items 13 “Wash or blow dry your hair,” 26 “Tingling” and 29 “Difficulty sleeping” were removed, as they had χ^2 probabilities at the 1% significance level in individual item fit analyses (Table 2). After removing items 13, 26 and 29, χ^2 fit residuals showed that item 2 “Writing” had a 5% significance level but removing the item 2 did not decrease the number of significant t-tests (indicator of unidimensionality). Furthermore, there were no statistically significant probabilities in item fit after removing item 2. Thus, item 2 was kept. Besides item 2, items 4, 8, 21, 27, 28 and 30 were potentially misfitting, as they had item fit residual over ±2.5. Further removal of items 28 and 27 resulted in a unidimensional scale (Table 3).

Table 3. Item reduction analysis monitor illustrating summary statistics for each step of adjustment.[AQ6]

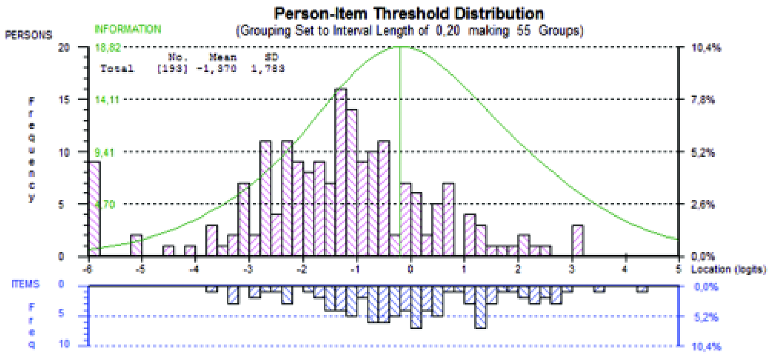
	Item location	Person location	Item fit residual	Person fit residual	χ^2 interaction				% of significant t-tests
Analysis	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD	Value	df	p	PSI	
Initial	0 (0.69)	–1.19 (1.58)	0.02 (2.29)	–0.24 (1.54)	169	60	0	0.95	27.08
Removed item 29	0 (0.70)	–1.23 (2.48)	0 (2.18)	–0.24 (1.49)	174	58	<0.001	0.95	9.73
Removed item 13	0 (0.70)	–1.19 (1.62)	0.02 (2.09)	–0.24 (1.48)	155	56	<0.001	0.95	8.11
Removed item 26	0 (0.73)	–1.24 (1.68)	0.03 (1.93)	–0.24 (1.43)	95	54	<0.001	0.95	7.77
Removed item 28	0 (0.75)	–1.28 (1.71)	0.00 (1.85)	–0.23 (1.37)	98	52	<0.001	0.95	7.57
Removed item 27	0 (0.77)	–1.32 (1.75)	0.02 (1.88)	–0.22 (1.32)	95	50	<0.001	0.95	4.89
Collapsed item 20 rc 2–4	**0 (0.79)**	**–1.35 (1.77)**	**0.02 (1.90)**	**–0.22 (1.34)**	**104**	**50**	**<0.001**	**0.95**	**4.66**
Collapsed item 21 rc 2–4	**0 (0.82)**	**–1.37 (1.78)**	**–0.04 (1.89)**	**–0.21 (1.31)**	**109**	**40**	**<0.001**	**0.95**	**4.15**

Testlet formation

As another option creating two testlets led to a unidimensional scale (χ^2 0.64, degrees of freedom 4, p = 0.96). The first testlet comprised of 25 items (1, 2, 4, 5, 7, 8, 11, 9, 10, 12, 13, 14, 15, 16, 18, 19, 20, 22, 24, 25, 26, 27, 28, 29 and 30). The second testlet consisted of the remaining five items (3, 6, 17, 21 and 23). In t-tests, 2.87% of estimates were significantly different. Each testlet explained 100% of the total non-error variance, supporting a one factor structure. Persons’ item fit did not exceed the threshold of ±2.5. PSI was 0.95.

Figure 1 illustrates the person–item distribution of the original version of the Finnish DASH providing good coverage and targeting for patients between logits –3 and 0.

Figure 1. Person-item distribution, good coverage for patients with ability between –3 and 0 logits.



In the original Finnish version of the DASH, seven items exhibited DIF ($p < 0.01$, Table 2). After Bonferroni correction, only item 5 ("Difficulty pushing open a heavy door") showed significant DIF by age.

Discussion

The Rasch analysis of the Finnish version of the DASH revealed multidimensionality, high levels of residual correlation between items, disordered item thresholds and partially poor item fit to the model.

The problem of multidimensionality mirrors findings from previous studies.[5,13] It has been suggested that the DASH could measure a combination of traits including pain, impact and function.[16] Unidimensionality can be achieved by item reduction, but this may remove clinically important items.[7,11] In the present study, a total of five items were removed, 13 ("Wash or blow-dry your hair"), 26 ("Tingling"), 27 ("Weakness"), 28 ("Stiffness") and 29 ("Difficulty sleeping"). The last four items describe symptoms the patient has experienced in the last week. These do not describe function, and thus it is understandable that they are a source of multidimensionality.

As another option to achieve unidimensionality, two testlets were formed. However, the items in the testlets did not seem to have a logical clinical relation to each other and do not lend themselves well to forming two separate subscales. The smaller of the two testlets comprised of items from the physical activities section of the DASH, items 3 ("Turn a key"), 6 ("Place an object on a shelf above your head"), 17 ("Recreational activities which require little effort"), 21 ("Sexual activities") and 23 ("Were you limited in your work or other daily activities").

Item thresholds are rating scale transition points. At these points, the person is as likely to answer in either category (1 or 2, 2 or 3, etc.). In the present analysis, item thresholds were disordered for items 20 ("Manage transport needs") and 21 ("Sexual activities"). Patients' answers to these items were bunched together implying that the patients are unable to differentiate between the response options.

The item fit was poor for items 26 ("Tingling") and 29 ("Difficulty sleeping"). Tingling has not fitted the Rasch model also in previous studies of the DASH and the QuickDASH questionnaires.[7,29,30] These papers have had between 3 and 16 items that have not fit the Rasch Model.[7,30] The present study population was a heterogeneous sample of patients with hand complaints. Although these items might fit the Rasch model for a single specific hand pathology, tingling and difficulties sleeping are non-specific symptoms that are present in many maladies that do not affect the function of the upper limb.

DIF arises when a subgroup of patients with the same underlying ability has a different probability of giving a certain answer. For example, young patients and old patients with the same level of disability may give different answers to an item. With uniform DIF, the difference in probability remains constant along different levels of ability. With non-uniform DIF, the subgroups have different probabilities at different levels of ability. The original Finnish DASH had seven items that exhibited DIF at $p < 0.01$ (Table 2). This suggests that women may report being less able to perform physical activities at the same level of disability. Similar findings were reported by Forget et al.[7] In the present study, however, only item 5 ("Difficulty opening a heavy door") exhibited significant DIF by age after Bonferroni correction ($p < 0.00057$).

Altogether 87 pairs of items exhibited residual correlation above 0.2. High levels of local dependency artificially inflate the reliability of the Finnish DASH. This might be reflected in the high PSI of 0.95 of the present analysis. In earlier papers, Baker et al. reported that 65 pairs of items had residual correlations exceeding 0.3 in patients after a cerebral stroke.[30] Franchignoni's Rasch analysis of the Italian version of the DASH showed residual correlation between 10 item pairs.[13] This means, according to the Rasch analysis, that many of the items may measure similar traits adding unnecessary bulk to the questionnaire without adding any new information. Even so, from a clinical perspective, similar items may bring additional information that is not registered by the Rasch model. For example, items 24 and 25 ("Arm, shoulder or hand pain" and "Arm, shoulder or hand pain when you performed any specific activity") had the highest residual correlation but could guide the clinician as to how much pain medication the patient needs.

For most patients in the present sample, the original Finnish DASH scale was able to provide good coverage and targeting. Item scaling was evenly distributed, and person ability estimate had a mean of –1.37 (Figure 1). There was a lack of items targeting low disability (gap in the distribution of items on the left). Earlier studies have partly reported the opposite. Braitmayer et al. reported poor discrimination for patients with low disability when analysing the DASH with a similar study population.[12] Forget et al. also found that the DASH had better coverage and targeting in patients with high disability in a sample made up of Dupuytren's contracture patients.[7]

The present patient sample represented a mixture of typical patients encountered at the hand surgery clinic (see Table 1). The results are thus generalizable mainly for these patients. One weakness of this study is its cross-sectional design, which did not allow the authors to further investigate the responsiveness or test–retest reliability of the DASH. In order to be effective in clinical use, the responsiveness of the DASH needs to be described for patients with hand complaints.

Conclusions

The Finnish version of the DASH seems to assess also traits other than disability. Two items did not fit the Rasch model and principal component analysis revealed multidimensionality. Because of multidimensionality, the clinician must be careful when interpreting changes in the score. The change in the score might not truly represent a change in the patients underlying traits of function or symptoms.

Although the Rasch analysis suggests that there is unnecessary bulk to the questionnaire (high residual correlation between items), similar items may still bring additional information to the clinician. Patients had problems differentiating between response options for two items (disordered item thresholds). There was minor response bias by age on one item.

Even with the aforementioned problems, the original Finnish DASH covers different levels of disability reasonably well, can distinguish between patients of varying levels of disability (good item coverage) and did not have a response bias by age.

Acknowledgements

The authors are thankful to the staff of Länsi-Pohja's Central Hospital surgery outpatient clinic for their support in conducting the current study.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

Ethical approval

Written informed consent was obtained from the patients for their anonymized information to be published in this article. The study was approved by the Ethics Committee of the Northern Ostrobothnia Hospital District, Finland (ETTKMK 34/2017).

Guarantor

JI.

Contributorship

JI and SH were equal contributors in this study. JI and JPR researched the literature and conceived the study. JI wrote the manuscript and analysed the data. SH gained ethical approval, recruited the patients and collected the data. JPR was involved in protocol development, analysed the data and supervised the article. JR, AH and JK reviewed and edited the article. All authors reviewed and approved the final version of the article.

ORCID iDs

J Ikonen https://orcid.org/0000-0002-8778-6573

Sina Hulkkonen ORCID: 000-0002-0555-047X

J Ryhänen https://orcid.org/0000-0002-4676-2831

Jussi P. Repo ORCID: 0000-0002-4688-5698 ~~Sina Hulkkonen ORCID: 000-0002-0555-047X Jorma Ryhänen ORCID: 0000-0002-4676-2831~~

Arja Häkkinen ORCID: 0000-0002-5779-1259

Jaro Karppinen ORCID: 0000-0002-2158-6042

References

1. MacDermid JC. Patient-reported outcomes: state-of-the-art hand surgery and future applications. Hand Clin 2014; 30: 293–304.

2. Hudak PL, Amadio PC and 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. Erratum in: *Am J Ind Med* 1996; 30: 372.

3. Beaton DE, Katz JN, Fossel AH, et al. 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: 128–146.

4. Gummesson C, Atroshi I and 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.

5. Prodinger B, Hammond A, Tennant A, et al. Revisiting the disabilities of the arm, shoulder and hand (DASH) and QuickDASH in rheumatoid arthritis. BMC Musculoskelet Disord 2019; 20: 41.

6. Cano SJ, Barrett LE, Zajicek JP, et al. Beyond the reach of traditional analyses: using Rasch to evaluate the DASH in people with multiple sclerosis. Mult Scler 2011; 17: 214–222.

7. Forget NJ, Jerosch-Herold C, Shepstone L, et al. Psychometric evaluation of the Disabilities of the Arm, Shoulder and Hand (DASH) with Dupuytren's contracture: validity evidence using Rasch modeling. BMC Musculoskelet Disord 2014; 15: 361.

8. Hacklin E, Timlin S, Madanat R, et al. DASH-kyselykaavakkeen suomentaminen ja kulttuuriadaptaatio. SOT 2009; 32: 252–254.

9. ~~Uimonen~~Uimonen MM, Hulkkonen SM, Ryhänen J, et al. Assessment of construct validity of the Finnish versions of the Disabilities of Arm, Shoulder and Hand Instrument and the Michigan Hand Outcomes Questionnaire [published online ahead of print, 2019 Aug 31]. *J Hand Ther*. ~~MM, Hulkkonen SM, Ryhänen J, et al. Assessment of construct validity of the Finnish versions of the Disabilities of Arm, Shoulder and Hand Instrument and the Michigan Hand Outcomes Questionnaire. *J Hand Ther* 2019~~[AQ4]

10. Tennant A and 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: 1358–1362.

11. Lundgren Nilsson Å and Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM™) revisited. J Rehabil Med 2011; 43: 884–891.

12. Braitmayer K, Dereskewitz C, Oberhauser C, et al. 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. Patient 2017; 10: 367–376.

13. 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.

14. Hung M, Voss MW, Bounsanga J, et al. Psychometrics of the patient-reported outcomes measurement information system physical function instrument administered by computerized adaptive testing and the Disabilities of Arm, Shoulder and Hand in the orthopedic elbow patient population. J Shoulder Elbow Surg 2018; 27: 515–522.

15. Beaton DE, Wright JG and 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.

16. Dalton E, Lannin NA, Laver K, et al. Validity, reliability and ease of use of the disabilities of arm, shoulder and hand questionnaire in adults following stroke. Disabil Rehabil 2017; 39: 2504–2511.

17. Lehman LA, Woodbury M and Velozo CA. Examination of the factor structure of the Disabilities of the Arm, Shoulder, and Hand questionnaire. Am J Occup Ther 2011; 65: 169–178.

18. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago, IL: MESA Press, 1993.

19. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. J Rehabil Med 2003; 35: 105–115.

20. Hagquist C, Bruce M and Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. Int J Nurs Stud 2009; 3: 380–393.

21. Hobart J and Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009; 13: 1–177.

22. Andrich D. Rating scales and Rasch measurement. Expert Rev Pharmacoecon Outcomes Res 2011; 11: 571–585.

23. Wright BD and Stone MH. Reliability and separation. In: *Measurement essentials*. Wilmington, DE: Wide Range, INC., 1999, pp.157–165.[AQ5]

24. Anderson EB. A goodness-of-fit test for the Rasch model. Psychometrika 1973; 38: 123–140.

25. Linacre JM and Wright BD. (Dichotomous Mean-square) Chi-square fit statistics. Rasch Meas Trans 1994; 8: 360–361.

26. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3: 205–31.

27. Smith RM. Person response maps for rating scales. Rasch Meas Trans 1994; 8: 372–373.

28. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010; 19: 539–549.

29. Franchignoni F, Ferriero G, Giordano A, et al. Psychometric properties of QuickDASH – a classical test theory and Rasch analysis study. Man Ther 2011; 16: 177–182.

30. Baker K, Barrett L, Playford ED, et al. Measuring arm function early after stroke: is the DASH good enough? J Neurol Neurosurg Psychiatry 2016; 87: 604–610.

General Queries

Query: GQ1: Please confirm that all author information, including names, affiliations, sequence, and contact details, is correct.

Query: GQ2: Please review the entire document for typographical errors, mathematical errors, and any other necessary corrections; check headings, tables, and figures.

Query: GQ3: Please confirm that the Funding and Conflict of Interest statements are accurate.

Query: GQ4: Please ensure that you have obtained and enclosed all necessary permissions for the reproduction of artistic works, (e.g. illustrations, photographs, charts, maps, other visual material, etc.) not owned by yourself. Please refer to your publishing agreement for further information.

Query: GQ5: Please note that this proof represents your final opportunity to review your article prior to publication, so please do send all of your changes now.

Author Queries

Query: AQ1 Please check whether '6 as 25' in the address is correct as given.
Response: as is short for asunto (apartment in finnish)

Query: AQ2 'World Medical Association 2013' looks like a reference citation. Kindly consider giving complete details in the reference list and provide the corresponding superscript number in the citation in sequential order.
Response: World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310(20):2191–2194.

Query: AQ3 Please check edit of punctuations in 'Items 24 "Arm, shoulder …'
Response: Ok

Query: AQ4 Please update ref. 9.
Response: Answered within text

Query: AQ5 Please provide the editor names in ref. 23.
Response: the copy that I have only mentions that the book was published by Wide Range Inc. They do not name an editor for the book in the foreword or in the title page

Query: AQ6 Two captions were given for Table 3. Kindly confirm that the one set here is correct.
Response: Ok

Comments

C1 Author: After Bonferroni correction, only item 5 "Difficulty opening a heavy door" exhibited DIF by age. The other items that ehibited had a less stringent p-value (0.01);