medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Manual ability in hand surgery patients: validation of the ABILHAND scale in four diagnostic groups

Ghady EL KHOURY MD[1,2], Olivier BARBIER MD[1], Xavier LIBOUTON MD,PhD[1], Jean-Louis THONNARD PhD[2], Philippe LEFEVRE PhD[2], Massimo PENTA PhD[2,3]

[1]Service d'orthopédie et traumatologie, Cliniques universitaires Saint-Luc, Avenue Hippocrate 10, 1200 Bruxelles, Belgique

[2]Institute of Neuroscience, Université catholique de Louvain, Avenue Mounier 53, 1200 Bruxelles, Belgique

[3]Arsalis SPRL, Glabais, Belgium

Funding: This work was supported by a grant from the FRIA (Fonds pour la formation à la recherche dans l'industrie et l'agriculture) and the Louvain Bionics (Incentive Grant 2018). The funding source had no role in study design, in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Conflicts of interest: The authors declare that they have no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Corresponding author: Ghady El Khoury, Service d'orthopédie et traumatologie, Cliniques universitaires Saint-Luc, Avenue Hippocrate 10, 1200 Bruxelles, Belgique

Email: ghady.elkhoury@uclouvain.be

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background: Patients treated in hand surgery (HS) belong to different demographic groups and have varying impairments related to different pathologies. HS outcomes are measured to assess treatment results, complication risks and intervention reliability. Quantitative analysis of functional recovery requires a unidimensional, linear scale.

Objective: To adapt the ABILHAND questionnaire through Rasch analysis for specific use in HS patients and to examine its validity.

Methods: A preliminary 90-item questionnaire was presented to 216 patients representing the diagnoses most frequently encountered in HS, including distal radius fracture (n=74), basal thumb arthritis (n=66), carpal tunnel syndrome (n=53), and heavy wrist surgery (n=23). Patients were assessed during the early recovery and in the late follow-up period (0-3 months, 3–6 months and >6 months), leading to a total of 305 assessments. They rated their perceived difficulty with queried activities as impossible, difficult, or easy. Responses were analyzed using the RUMM2030 software. Items were refined based on item-patient targeting, fit statistics, differential item functioning, local independence and item redundancy. Patients also completed the QuickDASH, 12-item Short Form Survey (SF-12) and a numerical pain scale.

Results: The rating scale Rasch model was used to select 23 mostly bimanual items on a 3-level scale, which constitute a unidimensional, linear measure of manual ability with good reliability across all included diagnostic groups (Person-Separation Index = 0.90). The resulting scale was found to be invariant across demographic and clinical subgroups and over time. ABILHAND-HS patient measures correlated significantly ($p<0.001$) with the QuickDASH ($r=-0.77$), SF-12 Physical

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Component Summary (r=0.56), SF-12 Mental Component Summary (r=0.31), and pain scale (r=-0.49).

Conclusion: ABILHAND-HS is a robust person-centered measure of manual ability in HS patients.

KEYWORDS: outcome assessment, hand, activities of daily living, surgery, patient reported outcome measures, Rasch analysis.

List of abbreviations

BTA: Basal Thumb Arthritis

CTQ: Carpal Tunnel Questionnaire

CTS: Carpal Tunnel Syndrome

DASH: Disability of Arm, Shoulder and Hand questionnaire

DIF: Differential Item Functioning

DRF: Distal Radius Fracture

HS: hand surgery

HWS: Heavy Wrist Surgery

MAM: Manual Ability Measure

MCS: Mental Component Summary

PCS: Physical Component Summary

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

PRWE: Patient Rated Wrist Evaluation

PSI: Person Separation Index

SF-12: 12-item Short Form Survey

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

In hand surgery (HS), as in other medical specialties, outcome evaluations are needed to assess the effectiveness and reliability of the intervention, as well as to reinforce patient education regarding the risks and outcomes of the procedure and, potentially, to justify therapeutic practices to payers.[1] Physician-documented reports of HS outcomes based on clinical examination and imaging should be complemented with patient reported outcomes assessed by questionnaires designed to capture patients' perspectives with respect to the impact of their conditions and interventions on their daily lives.[2–4]

Current views of health and disability have been shaped by the World Health Organization's International Classification of Functioning, Disability, and Health[5], which parses disease consequences into three domains: impairment of anatomical structures (e.g. bones, muscles, ligaments) or body functions (e.g. motor skills, sensitivity), activity limitations (e.g. manual activities), and participation restrictions (e.g. in hobbies and work). Although impairment measurements such as imaging can provide clues regarding functional prognosis, it does not provide good information about performance in everyday life, especially of the hands, which are important for a great variety of activities.[6–8] For example, demonstration of a bone fracture union is insufficient to determine whether a patient is capable of resuming their usual activities or returning to work.[9–12]

The patient-reported questionnaires that have been most commonly used in HS[13] are the Disability of the Arm, Shoulder and Hand questionnaire[14] (DASH), the Patient Rated Wrist Evaluation[15] (PRWE), and the Carpal Tunnel Questionnaire[16] (CTQ). Each of these questionnaires has been reported to have good psychometric properties, but each has a particular focus on its

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

own area(s) of disablement. The DASH assesses body functions, activities, and participation[17] and can be divided into 3 subscales based on dimensionality[18]. Meanwhile, the PRWE is specific to the wrist joint and the CTQ is specific to carpal tunnel syndrome (CTS). Interpretation of total scores on multi-dimensional instruments can be less than straightforward given that patients can show simultaneous improvement in one domain with deterioration in another.[19,20] Assessment of functional recovery on a unidimensional[21] and linear[20] scale would allow for quantitative comparisons of ability among different patients and treatments. Such a scale can be developed with state-of-the-art psychometric methods, such as the Rasch model.[22,23]

The ABILHAND questionnaire is a Rasch-model built measure of manual ability[24] that provides an invariant linear scale and allows for quantitative comparisons of manual ability between patients and over time. The scale has been validated in populations with rheumatoid arthritis,[25] chronic stroke,[8] pediatric cerebral palsy,[26] systemic sclerosis[27] and neuromuscular diseases.[28] These previous validations have shown that the difficulty of most manual activities was diagnosis-dependent.[29] Therefore, the objective of this work was to adapt the ABILHAND scale to the most frequent diagnoses treated in HS.

Methods

Questionnaire adaptation to HS patients

The ABILHAND is a measure of manual ability that assesses one's ability to manage daily activities requiring upper limb use, regardless of strategy[24]. To develop a HS-adapted ABILHAND, a preliminary item list was compiled from previous versions of the ABILHAND questionnaire, the DASH, PRWE, CTQ, and Michigan Hand Questionnaire items[30] together with some new items. This

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

pool of items was submitted to nine HS experts (hand surgeons, physical medicine and rehabilitation physicians, physical therapists, and occupational therapists), who were asked to assess each item's relevance and propose additional ones. A final list of 90 items constituted the experimental ABILHAND-HS questionnaire.

Patients

A convenience sample of 216 patients was recruited from the HS consultation center at Cliniques Universitaires Saint-Luc, Belgium representing the following four diagnostic categories: CTS, distal radius fracture (DRF), basal thumb arthritis (BTA), and heavy wrist surgery (HWS, including 1st row carpectomy and partial or total wrist arthrodesis). The inclusion criteria for patients were being >18 years old and being able to read and understand French. The exclusion criteria included any comorbidity that may impede manual ability substantially (e.g. tremor, paralysis, or active rheumatologic disease) and any mental or cognitive dysfunction. The patient characteristics are summarized in Table 1. Patients provided written informed consent to participate. This study was approved by the ethical committee of Cliniques Universitaires Saint-Luc-Université catholique de Louvain.

Procedures

The French-language experimental ABILHAND-HS items were presented in five random orders to avoid a systematic item sequence bias. Patients were asked to indicate their perceived difficulty associated with completing the activities without technical or human assistance, independent of the hand used to perform the activity on a three-level scale: impossible (0), difficult (1), or easy (2).[8] Activities not attempted during the last week were treated as missing responses. Patients

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

also completed the QuickDASH[31], 12-item Short Form Survey (SF-12) questionnaire[32] and a 10-level numerical pain scale, for external validation purposes.

Patients were first assessed as soon as they presented to their hand surgery consultation appointments and had experienced manual activities in their own environment: after hand surgery and cast removal for DRF, BTA and HWS and at the first consultation for non-operated CTS and BTA. For the first assessment, patients were interviewed by the principal investigator in order to ensure clarity, obtain feedback from participants, and make sure instructions are properly followed. Patients were also asked to suggest additional items they felt the questionnaire was missing. Follow-up assessments were completed in our consulting office or returned by mail, leading to a total of 305 completed assessments.

Rasch analysis

The 90-item experimental ABILHAND-HS questionnaire responses were analyzed using the Rasch model in RUMM2030 software (RUMM Laboratory Pty Ltd., Perth, Australia). The Rasch model[23], a prescriptive model, requires that specified response probabilities depend on only item difficulty and patient ability. Polytomous datasets with thresholds between successive response categories can be analyzed with either a rating scale model that constrains all threshold locations to be equal across items[33] or a partial credit model that allows threshold locations to vary across items.[34] Patient abilities and item difficulties are located along a common linear, unidimensional continuum that defines the latent variable of interest (i.e. manual ability). The locations are expressed in logits, calculated as the logarithm of the pass/fail probability ratio of an item or threshold. The logit locations were converted into centiles to facilitate clinical interpretation on

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

a linear scale ranging from 0% (least ability) to 100% (greatest ability).[35] Expected responses, determined based on the patient and item locations, were compared to the responses actually reported to compute residual and fit statistics, which were then used to assess the scale's unidimensionality.[36] A good fit of the data with the model affirms invariant locations along the continuum and indicates that the measure can be used to compare manual ability across patients and diagnoses.

Item selection

From the experimental version of the questionnaire, the ABILHAND-HS was refined through successive analyses of 305 assessments with the goal of selecting items that define a unidimensional and clinically relevant scale of manual ability. P values < 0.05 were considered significant for each of the following analysis steps:

1) *Item-patient targeting.* Based on examination of patient distributions and item locations, items that showed a floor effect (too easy) or did not target the patients sample ability were removed.
2) *Rating scale*. Items with disordered thresholds and items with thresholds that were too narrow (<1.4 logits) or too wide (>5 logits) were removed before applying the rating scale model.[37]
3) *Unidimensionality*. Only items that delineated a common manual ability construct according to the following four criteria were retained: (1) standardized residuals obtained over three class intervals had to be within ±2.5 with a non-significant χ^2;[36] (2) no observable major differential item functioning[38] (DIF), uniform or non-uniform, shown by a 2-way analysis of

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

variance of the residuals with Bonferroni correction,[39] according to gender (male vs. female), age (above vs. below the median age of 63 years), pathology (CTS vs. DRF vs. BTA vs. HWS), involved hand (dominant vs. non-dominant), level of education (basic vs. superior), and follow-up (0-3 months vs. 3–6 months vs. >6 months); (3) overall fit of the response set based on a non-significant item-trait interaction χ^2;[36] and (4) statistically similar patient locations, according to paired t-tests, calculated with items that loaded either positively or negatively on the first residuals principal component.[40–42]

4) *Local independence.* When items were found to be querying redundant content,[43] demonstrated by a residual correlation > 0.3, the item with the poorer fit statistic was deleted.[44]

5) *Item redundancy*. To shorten the scale, when two or more items had similar locations on the continuum, the one with the best fit was retained.

Scale reliability

The Person-Separation Index (PSI), i.e. the proportion of total variance (including error) that is attributed to patient location variance, was used to determine the ABILHAND-HS scale's reliability and its degree of precision with the dataset, and thus how many statistically different ability strata can be distinguished along the scale.[45]

Construct validity

The construct validity of the ABILHAND-HS was examined with Pearson correlational analysis for association with age, t-tests for associations with gender and involved hand, and an analysis of variance for diagnosis. The relationships of the ABILHAND-HS with the QuickDASH scale, the numerical pain scale, the SF-12 Physical Component Summary (PCS), and the SF-12 Mental

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Component Summary (MCS) were assessed with Spearman correlation coefficients. Patient perceptions were compared between ABILHAND-HS and QuickDASH items by adding the six QuickDASH activity items to the anchored data matrix. The locations of similar items were then compared between the scales.

Results

Item selection for the ABILHAND-HS scale

Successive analyses led to the selection of 23 items defining a unidimensional manual ability scale in HS. Of the 90 experimental items, 34 were removed because they were too easy (e.g. 'Drinking a glass of water'), 3 items had too-narrow thresholds (e.g. 'Using a touch screen'), 4 items were misfitting (e.g. 'Carrying a shopping bag'), and 26 items had a location redundant with another better fitting item (e.g. 'Peeling onions' was deleted in favor of 'Peeling potatoes with a knife').

Metric properties

The calibration obtained for the 23 mostly bimanual activities retained for ABILHAND-HS is reported in Table 2 in descending difficulty order. The standardized residuals obtained matched the expected standard normal distribution for items [mean (SD), -0.30 (0.99)] and for patients [0.31 (0.97)], indicating that the ABILHAND-HS scale is globally unidimensional. An invariant item location was obtained for more- and less- able patients as shown by a nonsignificant item-trait interaction (χ^2 = 57.76, 46 df, p = 0.11). An invariant patient ability was obtained with items with different content as shown by a non significant t-test when using items that loaded positively or negatively on the first principal residual component (t = 1.24, 304 df, p = 0.22).

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Analysis of DIF of the ABILHAND-HS with six criteria yielded only four instances of uniform DIF among the 23 items (Table 3). A small magnitude DIF was revealed among diagnoses (Fig. 1) with no substantial impact on scale invariance, as evidenced by a good overall fit. No DIF was observed between the first and last assessments, showing satisfactory invariance to support the scale follow-up stability. Likewise, an intraclass correlation coefficient across the first and last assessments was 0.94, indicating excellent item-difficulty-hierarchy consistency and providing confidence for data pooling over different time points.[46] The PSI in this sample was equal to 0.90, indicating the distinguishability of four strata of manual ability.[45]

Scale description

The ABILHAND-HS structure and targeting of HS patients are illustrated in Figure 2, showing an average patients' manual ability of 1.17 logits (SD = 1.85 logits; i.e. 58 (15) centiles). The three response categories were well distinguished in HS patients, with an inter-threshold distance of 2.93 logits (24 centiles), indicating that, regardless of patient ability, rating an item as 'easy' is about 20 (i.e. $e^{2.93}$=18.7) times more difficult than rating it as 'impossible'. Although the threshold distribution (range, -4.15 to 5 logits) was well targeted to the range of patient abilities, the patients' ability levels skewed high, indicating that the scale could measure patients that are more severely disabled than in this sample.

Construct validity

An effect of gender on ABILHAND-HS manual ability measures was observed, with men [mean (SD) of 1.88 (2.39) logits] reporting a significantly higher mean manual ability than women [1.32 (1.93) logits; 164 df, t = -2.0, p = 0.04]. Manual ability was not found to be significantly associated

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

with age (R = -0.04; p = 0.47), the hand involved (t = 0.96; 303 df; p = 0.37), or the patient's diagnosis (F = 1.92; 3 df; p = 0.12). Although variance across diagnosis groups was not significant, we did observe a broad spectrum of manual ability. Patients with CTS reported the highest manual ability [1.9 (2.0) logits], followed by patients with BTA [1.5 (2.3) logits], DRF [1.4 (2.1) logits], and HWS [0.9 (1.8) logits].

The relationships between ABILHAND-HS measures and scores obtained with other instruments are shown in Figure 3. Briefly, ABILHAND-HS correlated strongly with QuickDASH scores, moderately with SF-12 PCS scores and pain scale scores, and weakly with SF-12 MCS scores. We observed substantial similarity with respect to manual ability scale locations between the ABLHAND-HS and QuickDASH activity items (Fig. 4).

Discussion

Here, we report the adaptation and validation of an ABILHAND-HS questionnaire for use with HS patients. Impairments present in our study cohort included weakness (e.g. following DRF), loss of sensation (e.g. in CTS), and stiffness (e.g. in BTA), with some patients presenting with a combination of these impairments. The ABILHAND-HS was constructed to measure manual ability on a common, linear, and unidimensional scale wherein the 23 activities retained delineate an invariant item difficulty hierarchy independent of patient diagnosis. All ABILHAND-HS activities with the exception of one involve both hands and, consistent with our clinical experience, the most difficult ones require high levels of force. Of the experimental 90 items, those that could be interpreted in different ways, for instance using the injured or uninjured hand, were misfitting and thus omitted (e.g. 'carrying a shopping bag'). The sample size[47] and fit statistics for the 23

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

retained items support the item hierarchy invariance across the latent trait.[48] A few instances of minor DIF were retained to maintain the scale's construct validity.[49] The resulting scale is well targeted to the studied HS population, despite a small persistent ceiling effect, most likely due to missing responses for the most difficult activities. This observation of apparent ceiling effect involves 7.9% (24/305) of the records, which is well below the maximum recommended allowance of 15%.[50]

Although reliability indices should be compared with caution across potentially different study conditions, it is noteworthy that the PSI obtained for the ABILHAND-HS (0.90) was higher than prior values obtained for the activities subscales of the PRWE[51] (0.78 and 0.81 for the usual and specific activities subscales, respectively, in DRF patients), for the Patient-Rated Wrist and Hand Evaluation[52] (0.83 in HS patients), for the QuickDASH scale[53] (0.84 in patients with various upper limb dysfunctions) and for the Manual Ability Measure[54] (MAM-16; 0.83 for HS patients), while being equal to that for the DASH manual functioning subscale.[18] PSI values reflect sensitivity to clinical evolution over time, with greater values indicating a greater number of distinguishable ability strata. We obtained person separation among patients using three response levels (impossible, difficult, and easy), consistent with previous studies showing patients unable to discriminate more than three levels of difficulty (ABILHAND,[8] DASH activity items[18] and QuickDASH activity items[53]).

Accurate communication of scale administration instructions is critical for targeting patient manual ability as defined by the ABILHAND-HS. Generally, patients focus on their ability to perform the queried activities with their injured hand; likewise, the PRWE explores use of the affected hand explicitly.[15] The ABILHAND-HS, like the QuickDASH, is oriented towards real daily

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

life behaviors and is intended to be independent of the limb(s) or strategy used and unbiased by activities that are never performed with the affected hand or avoided during recovery.[8] Our findings of stable item calibrations and lack of DIF across the assessments indicate that the ABILHAND-HS can be used confidently to assess the patient recovery at different time points during follow-up. Moreover, the stability of items hierarchy between the first and last evaluation indicate that the results were not influenced by the method of administration (interview with the investigator versus self-reported).

ABILHAND-HS construct validation results fit well with our clinical observations. The patients with the highest manual ability scores on the ABILHAND-HS also had the highest SF-12 PCS and SF-12 MCS scores as well as the lowest QuickDASH and numerical pain scale scores. Correlations of ABILHAND-HS with other instruments, including the QuickDASH, SF-12 and a pain scale are also consistent with prior findings suggesting that generic instruments are less sensitive than specific ones.[55] The present ABILHAND-HS manual ability scores were not related significantly to age, consistent with other versions of the ABILHAND.[8,25,28] Our findings of a small, but significant gender effect, with men tending to report a higher manual ability than women (mean difference, 0.56 logits), varied across HS diagnoses but, generally, were consistent with previous reports in patients with DRFs and wrist arthrodesis.[56–58] The construct validity of the ABILHAND-HS was further supported by our confirmation of a similar item difficulty hierarchy for QuickDASH items in our patients sample. Notably, those ABILHAND-HS activities that require a great amount of force (e.g. 'Opening a screw-topped jar') have been reported to likewise be among the most difficult items in the DASH and QuickDASH[18,53] and in the MAM-16[54].

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The ABILHAND-HS, developed using Rasch methodology, has several advantages over questionnaires developed using classical test theory. Firstly, the ABILHAND-HS can tolerate missing responses, which enables it to remain valid even in patients who scarcely perform some of the queried activities. Secondly, the ability to analyze response patterns can identify those patients whose responses do not fit the model due to random or careless answers, a particular injury or comorbidities. Finally, the high precision of the ABILHAND-HS items minimizes the need for interpretation, thereby allowing more reliable comparisons between patients (e.g. recreational activities involving force or impact are broken down into the items 'doing push-ups', 'practicing a racket sport').

Conclusion

ABILHAND-HS was demonstrated to be a successful adaptation for application in HS patients. The resulting scale was shown to be a valid, patient-oriented, clinically meaningful and precise instrument. It targets commonly performed manual activities and allows stable and linear measurement of manual ability over multiple time points in patients treated for DRF, BTA, CTS, or HWS. The scale reveals unexpected responses that may provide clues regarding the patient's clinical state, as summarized at www.rehab-scales.org. Future research should include more patients with HWS, as well as other diagnoses such as tendinopathies and ligamentous injuries, and an assessment of scale responsiveness.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Dubert T. Outcome measurements in hand and upper limb surgery. Chirurgie de la Main. 2014;33:235–46.

2. Berkanovic E, Hurwicz M-L, Lachenbruch PA. Concordant and discrepant views of patients' physical functioning. Arthritis & Rheumatism. 1995;8:94–101.

3. Hewlett SA. Patients and Clinicians Have Different Perspectives on Outcomes in Arthritis. J Rheumatol. 2003;30:877–9.

4. Ziebland S, Fitzpatrick R, Jenkinson C. Tacit models of disability underlying health status instruments. Social science & medicine. 1993;37:69–75.

5. World Health Organization, editor. International classification of functioning, disability and health: ICF. Geneva: World Health Organization; 2001.

6. Barbier O, Penta M, Thonnard J-L. Outcome evaluation of the hand and wrist according to the International Classification of Functioning, Disability, and Health. Hand Clinics. 2003;19:371–8.

7. Bobos P, Lalone EA, Grewal R, MacDermid JC. Do Impairments Predict Hand Dexterity After Distal Radius Fractures? A 6-Month Prospective Cohort Study. HAND. 2018;13:441–7.

8. Penta M, Tesio L, Arnould C, Zancan A, Thonnard J-L. The ABILHAND Questionnaire as a Measure of Manual Ability in Chronic Stroke Patients: Rasch-Based Validation and Relationship to Upper Limb Impairment. Stroke. 2001;32:1627–34.

9. Giladi AM, Chung KC. Measuring Outcomes in Hand Surgery. Clinics in Plastic Surgery. 2013;40:313–22.

10. Jaremko JL, Lambert RGW, Rowe BH, Johnson JA, Majumdar SR. Do radiographic indices of distal radius fracture reduction predict outcomes in older adults receiving conservative treatment? Clinical Radiology. 2007;62:65–72.

11. Synn AJ, Makhni EC, Makhni MC, Rozental TD, Day CS. Distal Radius Fractures in Older Patients: Is Anatomic Reduction Necessary? Clin Orthop Relat Res. 2009;467:1612–20.

12. Young BT, Rayan GM. Outcome following nonoperative treatment of displaced distal radius fractures in low-demand patients older than 60 years. The Journal of Hand Surgery. 2000;25:19–28.

13. Changulani M, Okonkwo U, Keswani T, Kalairajah Y. Outcome evaluation measures for wrist and hand – which one to choose? International Orthopaedics. 2008;32:1–6.

14. 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.

15. MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. J Orthop Trauma. 1998;12:577–86.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

16. 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. JBJS. 1993;75:1585–1592.

17. Coenen M, Kus S, Rudolf K-D, Müller 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? Journal of Hand Therapy. 2013;26:332–42.

18. 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. Archives of Physical Medicine and Rehabilitation. 2010;91:1370–7.

19. Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. Archives of physical medicine and rehabilitation. 1989;70:308–312.

20. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. Arch Phys Med Rehabil. 1989;70:857–60.

21. Thurstone L. The measurement of values. Chicago: The University of Chicago Press; 1959.

22. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: Time to end malpractice? Journal of Rehabilitation Medicine. 2012;44:97–8.

23. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago, IL: The University of Chicago Press; 1980.

24. Penta M, Thonnard J-L, Tesio L. ABILHAND: a Rasch-built measure of manual ability. Archives of physical medicine and rehabilitation. 1998;79:1038–1042.

25. Durez P, Fraselle V, Houssiau F, Thonnard J, Nielens H, Penta M. Validation of the ABILHAND questionnaire as a measure of manual ability in patients with rheumatoid arthritis. Ann Rheum Dis. 2007;66:1098–105.

26. Arnould C, Penta M, Renders A, Thonnard J-L. ABILHAND-Kids: a measure of manual ability in children with cerebral palsy. Neurology. 2004;63:1045–52.

27. Vanthuyne M, Smith V, Arat S, Westhovens R, de Keyser F, Houssiau FA, et al. Validation of a manual ability questionnaire in patients with systemic sclerosis. Arthritis & Rheumatism. 2009;61:695–703.

28. Vandervelde L, Van den Bergh PYK, Penta M, Thonnard JL. Validation of the ABILHAND questionnaire to measure manual ability in children and adults with neuromuscular disorders. Journal of Neurology, Neurosurgery & Psychiatry. 2010;81:506–12.

29. Arnould C, Vandervelde L, Batcho CS, Penta M, Thonnard J-L. Can manual ability be measured with a generic ABILHAND scale? A cross-sectional study conducted on six diagnostic groups. BMJ open. 2012;2:e001807.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

30. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. The Journal of Hand Surgery. 1998;23:575–87.

31. Beaton DE, Wright JG, Katz JN. Development of the QuickDASH: Comparison of Three Item-Reduction Approaches. Journal of Bone and Joint Surgery (American). 2005;87A:9.

32. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34:220–33.

33. Andrich D. A rating formulation for ordered response categories. Psychometrika. 1978;43:561–73.

34. Masters GN. A rasch model for partial credit scoring. Psychometrika. 1982;47:149–74.

35. van Nes SI, Vanhoutte EK, van Doorn PA, Hermans M, Bakkers M, Kuitwaard K, et al. Rasch-built Overall Disability Scale (R-ODS) for immune-mediated peripheral neuropathies. Neurology. 2011;76:337–45.

36. Andrich D, Sheridan B, Luo G. Polytomous data. In: Interpreting RUMM2030: part II. Western Australia: Perth; 2013.

37. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas. 2002;3:85–106.

38. Holland PW, Wainer H. Differential item functioning. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc; 1993.

39. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt. 2014;34:502–8.

40. Linacre JM. Detecting multidimensionality: which residual data-type works best? J Outcome Meas. 1998;2:266–83.

41. 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;46:1–18.

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

43. Wright BD. Local dependency, correlations and principal components. Rasch Meas Trans 1996;10:509–11.

44. Ramp M, Khan F, Misajon RA, Pallant JF. Rasch analysis of the Multiple Sclerosis Impact Scale (MSIS-29). Health Qual Life Outcomes. 2009;7:58.

45. Fisher W. Reliability, Separation, Strata Statistics. Rasch Meas Trans. 1992;6:238.

46. Chang W-C, Chan C. Rasch analysis for outcomes measures: some methodological considerations. Archives of Physical Medicine and Rehabilitation. 1995;76:934–9.

47. Hagell P, Westergren A. Sample Size and Statistical Conclusions from Tests of Fit to the Rasch Model According to the Rasch Unidimensional Measurement Model (Rumm) Program in Health Outcome Measurement. J Appl Meas. 2016;17:416–31.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

48. 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 & Rheumatism. 2007;57:1358–62.

49. Hagquist C, Andrich D. Recent advances in analysis of differential item functioning in health research using the Rasch model. Health and Quality of Life Outcomes. 2017;15:181.

50. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Quality of Life Research. 1995;4:293–307.

51. Esakki S, MacDermid JC, Vincent JI, Packham TL, Walton D, Grewal R. Rasch analysis of the patient-rated wrist evaluation questionnaire. Arch. Physiother. 2018;8:5.

52. Packham T, MacDermid JC. Measurement properties of the Patient-Rated Wrist and Hand Evaluation: Rasch analysis of responses from a traumatic hand injury population. Journal of Hand Therapy. 2013;26:216–24.

53. Franchignoni F, Ferriero G, Giordano A, Sartorio F, Vercelli S, Brigatti E. Psychometric properties of QuickDASH – A classical test theory and Rasch analysis study. Manual Therapy. 2011;16:177–82.

54. Chen CC, Granger CV, Peimer CA, Moy OJ, Wald S. Manual Ability Measure (MAM-16): a preliminary report on a new patient-centred and task-oriented outcome measure of hand function. Journal of Hand Surgery. 2005;30:207–216.

55. Aktekin LA, Eser F, Başkan BM, Sivas F, Malhan S, Öksüz E, et al. Disability of Arm Shoulder and Hand Questionnaire in rheumatoid arthritis patients: relationship with disease activity, HAQ, SF-36. Rheumatol Int. 2011;31:823–6.

56. Amorosa LF, Vitale MA, Brown S, Kaufmann RA. A Functional Outcomes Survey of Elderly Patients who Sustained Distal Radius Fractures. Hand (New York, N,Y.). 2011;6:260–7.

57. Cowie J, Anakwe R, McQueen M. Factors Associated with One-Year Outcome after Distal Radial Fracture Treatment. Journal of Orthopaedic Surgery. 2015;23:5.

58. Owen DH, Agius PA, Nair A, Perriman DM, Smith PN, Roberts CJ. Factors predictive of patient outcome following total wrist arthrodesis. The Bone & Joint Journal. 2016;98-B:647–53.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure legends

Figure 1. Differential item functioning (DIF) plots comparing the item difficulty hierarchy between subgroups. In each plot, the lines represent the 95% confidence interval of an ideal invariance between subgroups; the items are represented by the dots or by their letter if they display significant DIF. Item difficulty hierarchy comparison between first and last assessments. The most difficult items (dots) are plotted in the top right part of each plot. When comparing the item difficulty hierarchy between each diagnostic group relative to the whole sample, most of the ABILHAND-HS items lie within 95% confidence interval of the ideal invariance, indicating an invariant difficulty across diagnostic groups. When comparing the item difficulty hierarchy between the first and last assessment, all items fall within the 95% confidence interval of an ideal invariance, affirming invariance of item difficulties between the assessments.

Figure 2. Structure of the ABILHAND-HS scale. Top: distribution of manual ability measures for the whole sample expressed in logits (log of the pass/fail probability ratio) and centiles (fraction of the measurement range). Twenty-four patients (7.9%) were able to perform all 23 activities easily, and were thus identified as extreme patients. None of the participants reported that they could not perform any of the 23 activities. Middle: most probable patient response to each item based the patient manual ability and on the difficulty of the item's response category. The average item difficulty was set to 0 logits and the items are ordered from most (top) to least (bottom) difficult. The distance between thresholds (middle bar) is constant for all items (2.93 logits or 24 centiles). A patient with a manual ability measure of 0 logits would be expected to perform the first 3 activities easily, to have some difficulty with the following 17 activities, and to be unable to perform the 3 most difficult activities. A patient with a measure of 2.1 logits

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

should be able to perform all activities easily or with some difficulty. Bottom: conversion of ordinal raw scores into a linear continuum of manual ability for complete response sets. The raw scores ranged from 0 to 46 (sum of scores of 0–2 for 23 items). This curve is linear in its central (30th~70th percentile) range, with sigmoid flattening outside the central range, highlighting a non-linear relationship, especially at the extremities of the score range.

Figure 3. Correlations of ABILHAND-HS scores with QuickDASH, PCS, MCS, and numerical pain scale scores. Spearman correlation coefficients are indicated in the top right of each graph. All correlations were statistically significant ($p < 0.001$).

Figure 4. Comparison of difficulty levels (vertical axis) between similar items of the ABILHAND-HS (left) and QuickDASH (right) scales. QuickDASH item responses were added to the anchored data matrix of ABILHAND-HS responses to equate both measures.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Tables

Table 1. Sample characteristics (n = 216).

Characteristic	N (%)[a]
Gender	
Women	145 (67%)
Men	71 (33%)
Mean age (range), years	60.3 (19–93)
Education level	
Basic	109 (51%)
Postsecondary	107 (49%)
Work status	
Student	2 (1%)
Unemployed	22 (10%)
(Self-)Employed	83 (38%)
Retired	109 (51%)
Hand dominance	
Right	194 (90%)
Left	15 (7%)
Ambidextrous	7 (3%)
Involved dominant hand	
Yes	136 (63%)
No	80 (37%)
Diagnostic group	
Distal radius fracture (DRF)	74 (34%)
Basal thumb arthritis (BTA)	66 (31%)
Carpal tunnel syndrome (CTS)	53 (24%)
Heavy wrist surgery (HWS)	23 (11%)
Follow-up assessments (n = 305)	
0–3 months	132 (43%)
3–6 months	58 (19%)
>6 months	115 (38%)

[a]**Except where otherwise indicated**

medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

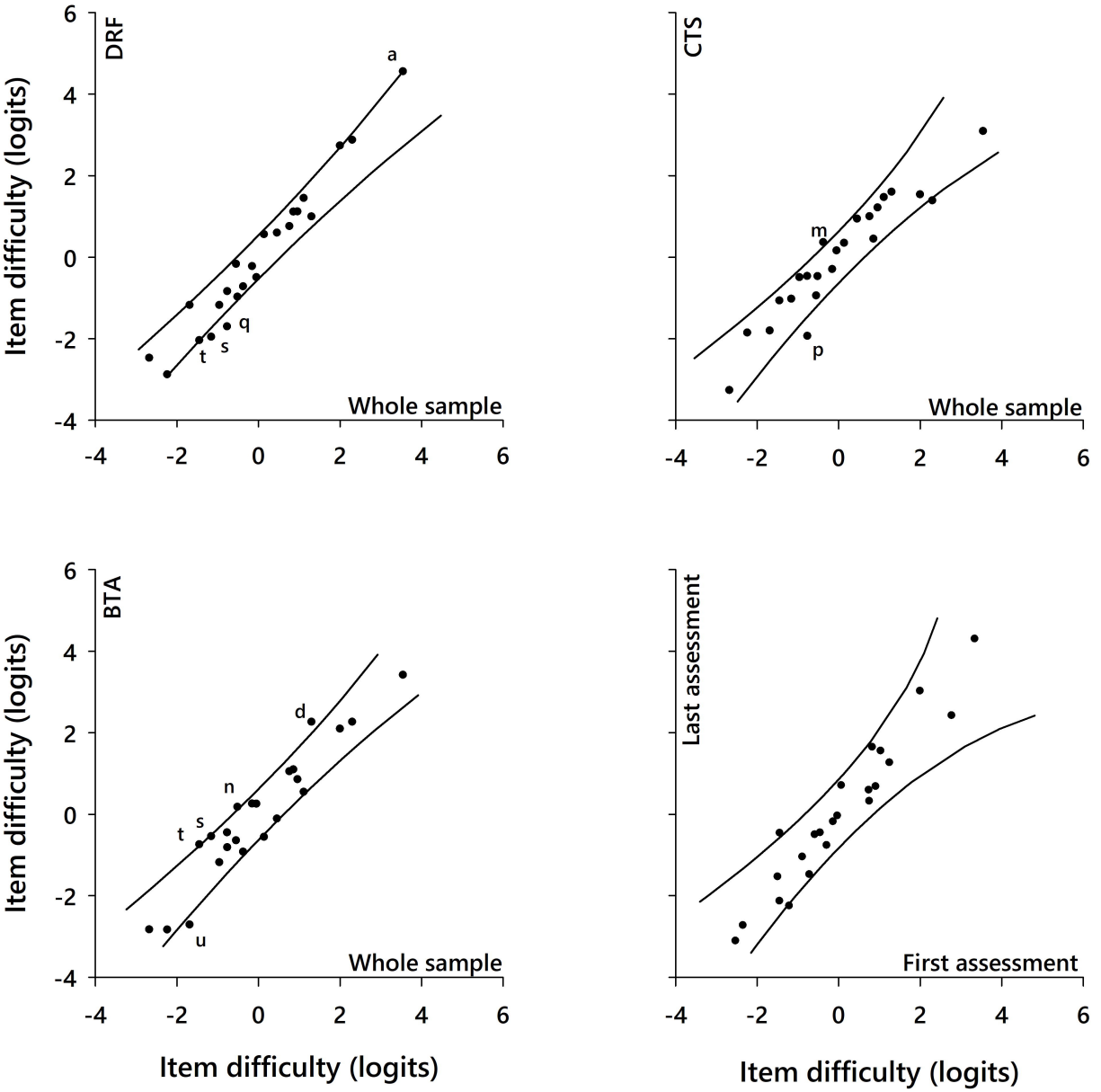
Table 2. Calibration of the 23 items of the ABILHAND-HS.

Item	Bi-manual	Difficulty logits (centiles)	SE logits	Residual z	Fit χ^2	*P*
a. Doing push-ups	x	3.54 (78)	0.21	0.61	0.46	0.79
b. Playing a racket sport	x	2.30 (68)	0.25	0.04	1.66	0.44
c. Cutting a hedge	x	2.00 (65)	0.18	0.05	3.83	0.15
d. Opening a screw-topped jar	x	1.30 (59)	0.12	0.48	6.19	0.05
e. Applauding vigorously	x	1.11 (58)	0.16	2.09	2.00	0.37
f. Lifting a full pan	x	0.96 (57)	0.12	-0.53	5.69	0.06
g. Wringing a towel	x	0.86 (56)	0.12	-2.00	0.14	0.93
h. Opening a can with a can opener	x	0.76 (55)	0.12	-1.83	0.49	0.78
i. Hammering a nail	x	0.45 (52)	0.16	-0.70	5.31	0.07
j. Shaking bed sheets	x	0.13 (50)	0.17	-1.95	3.49	0.17
k. Using a screwdriver	x	-0.05 (48)	0.14	-0.14	0.54	0.77
l. Peeling potatoes with a knife	x	-0.16 (47)	0.13	-0.75	5.75	0.06
m. Ironing	x	-0.38 (45)	0.15	-0.44	3.12	0.21
n. Taking the cap off a bottle	x	-0.52 (44)	0.13	-0.32	2.75	0.25
o. Cutting one's nails	x	-0.55 (44)	0.13	-0.05	0.33	0.85
p. Shuffling and dealing cards	x	-0.77 (42)	0.16	-1.21	6.18	0.05
q. Wiping windows		-0.77 (42)	0.15	0.01	1.29	0.52
r. Tying shoelaces	x	-0.96 (41)	0.14	-0.61	2.02	0.36
s. Tearing open a pack of chips	x	-1.16 (39)	0.15	1.70	1.31	0.52
t. Fastening the zipper of a jacket	x	-1.45 (37)	0.14	0.32	1.58	0.45
u. Turning a car steering wheel	x	-1.69 (35)	0.18	-0.26	0.78	0.68
v. Putting on gloves	x	-2.24 (30)	0.19	-0.80	1.99	0.37
w. Spreading butter on a slice of bread	x	-2.68 (26)	0.18	-0.60	0.86	0.65

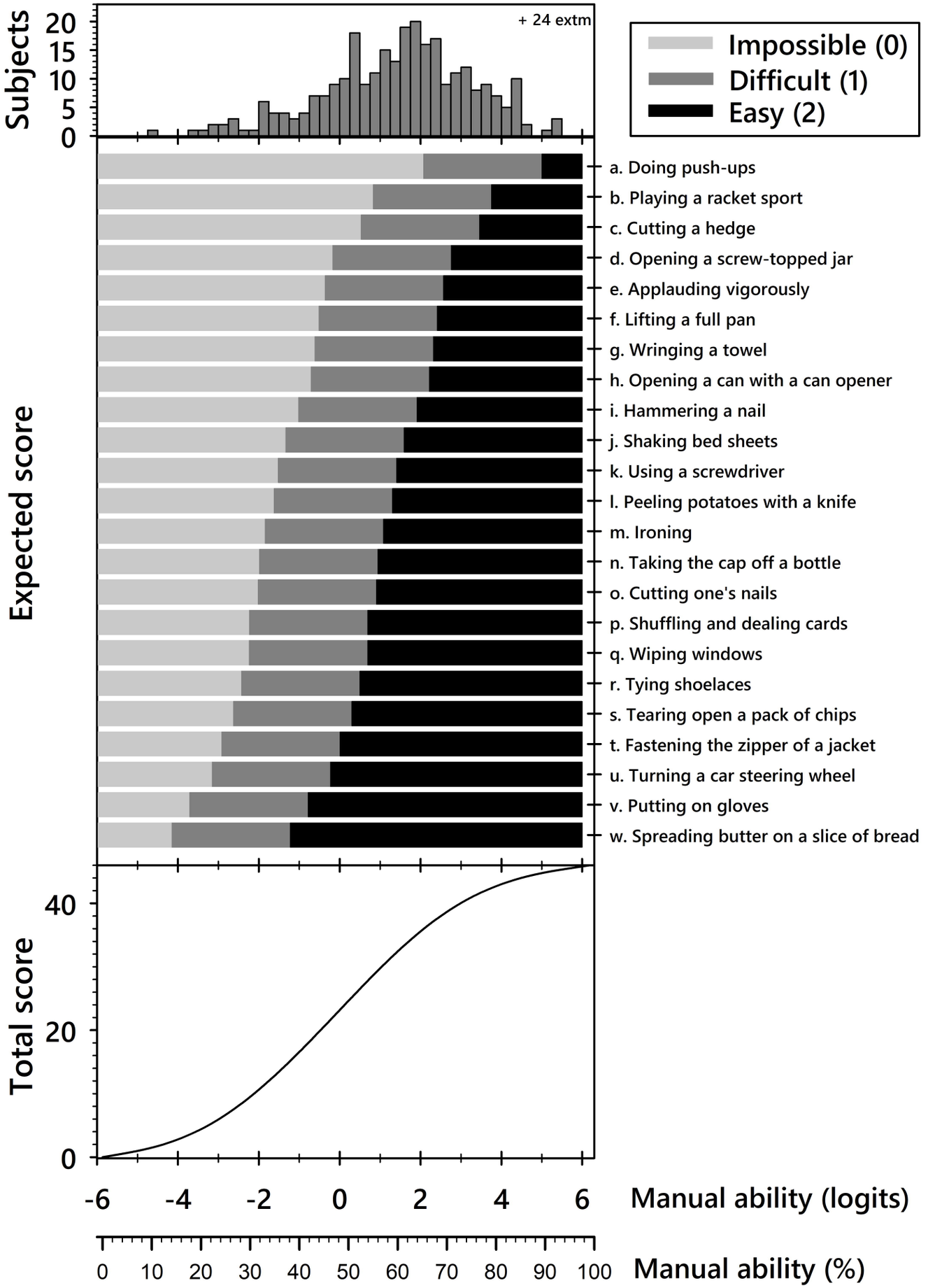
medRxiv preprint doi: https://doi.org/10.1101/2020.07.02.20144147; this version posted July 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

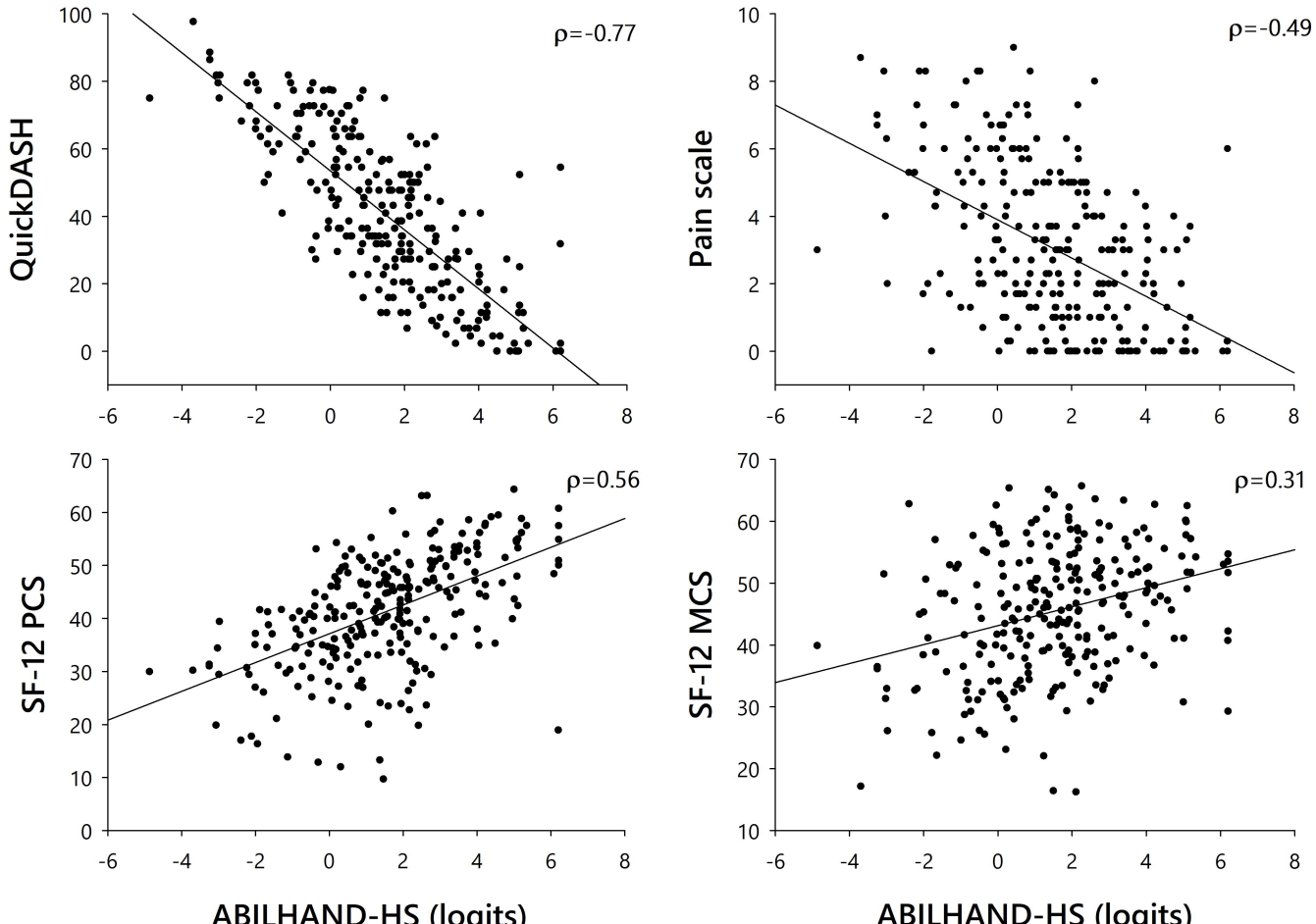
Table 3. Differential item functioning (DIF) summary.

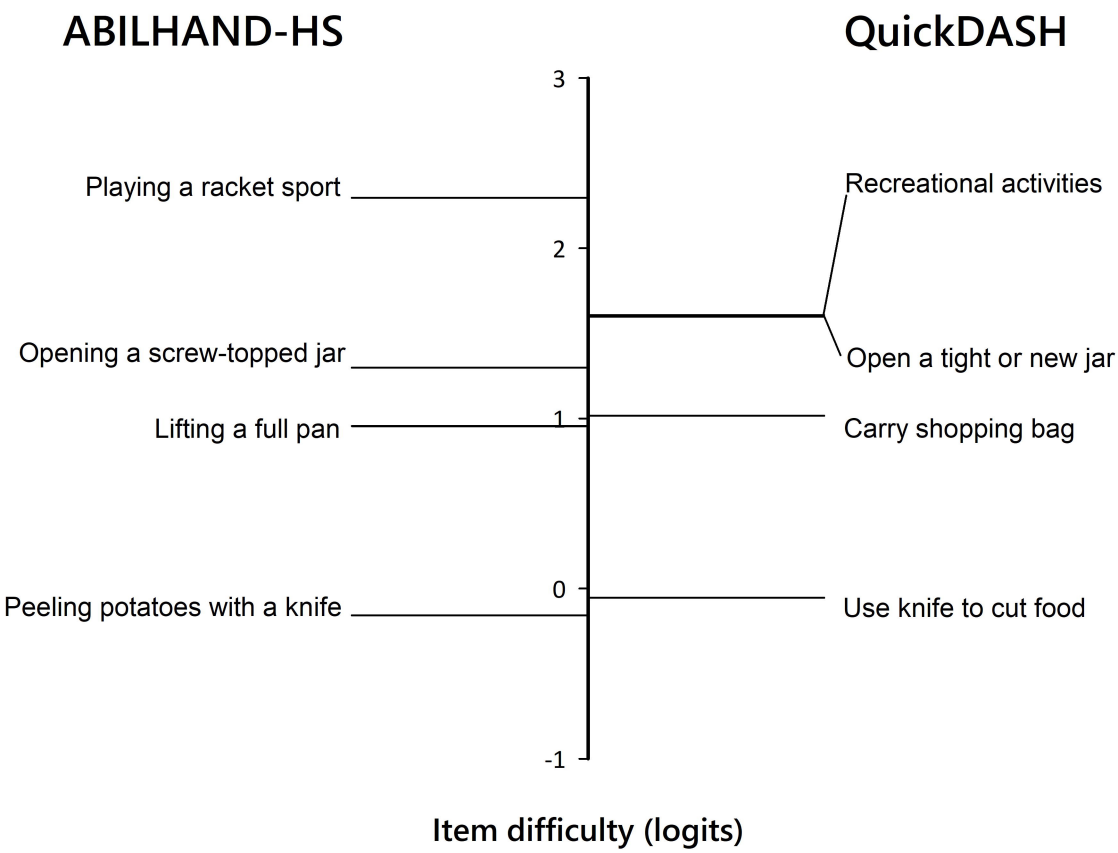
Label	Person factor	Magnitude (logits)	Type	Difficulty
Taking the cap off a bottle	Gender	1.36	Uniform	Women > Men
Opening a screw-topped jar	Gender	0.93	Uniform	Women > Men
Opening a screw-topped jar	Diagnosis	0.61	Uniform	BTA > CTS > DRF
Using a screwdriver	Involved hand	1.26	Uniform	Non-dominant > Dominant hand involved


DRF
a
q
t
s
Whole sample
CTS
m
p
Whole sample
BTA
d
n
t
s
u
Whole sample
Last assessment
First assessment
Item difficulty (logits)
Item difficulty (logits)
Item difficulty (logits)
Item difficulty (logits)
6
4
2
0
-2
-4
-4
-2
0
2
4
6

It is made available under a CC-BY-NC-ND 4.0 International license .




QuickDASH
ρ=-0.77
Pain scale
ρ=-0.49
SF-12 PCS
ρ=0.56
SF-12 MCS
ρ=0.31
ABILHAND-HS (logits)
ABILHAND-HS (logits)


ABILHAND-HS
QuickDASH
3
2
1
0
-1
Playing a racket sport
Opening a screw-topped jar
Lifting a full pan
Peeling potatoes with a knife
Recreational activities
Open a tight or new jar
Carry shopping bag
Use knife to cut food
Item difficulty (logits)