Creating a Clinically Useful Data Collection Form for the DASH Questionnaire

Leigh A. Lehman, Bhagwant S. Sindhu, Jessica Johnson, Craig A. Velozo

KEY WORDS
- activities of daily living
- data collection
- disability evaluation
- goals
- upper extremity

PURPOSE. We generated a clinically useful data collection form for the Disabilities of the Arm, Shoulder, and Hand (DASH) questionnaire. This data collection form is designed to aid in measuring change and goal setting.

METHOD. Rasch analysis was used to generate three data collection forms for constructs on the DASH (gross motor, fine motor, and symptoms; N = 960). A form was completed to represent the findings from 1 study participant. Admission and discharge data were illustrated for one of the three sections (gross motor activity). Possible goals were indicated on the admission form. The discharge form illustrates whether these goals have been achieved.

RESULTS. Figures illustrate the utility of the forms in observing functional change from admission to discharge and how the forms aid in goal setting.

CONCLUSION. Use of the data collection form has many positive implications. This type of form could aid in goal setting and treatment planning.


Occupational therapists can evaluate upper-extremity function using three different methods. First, they can choose from multiple performance measures (World Health Organization [WHO], 2002), which evaluate basic movements (e.g., elbow flexion/extension, wrist pronation/supination, and shoulder abduction/adduction), such as the Motor Assessment Scale (Carr, Shepherd, Nordholm, & Lynne, 1985), Fugl-Meyer Assessment (Fugl-Meyer, Jääskö, Leyman, Olsson, & Steglind, 1975), and Wolf Motor Function Test (Wolf et al., 2001). Second, therapists may use self-reports of daily functional ability of the arms and hands. These approaches include the familiar Disabilities of the Arm, Shoulder, and Hand (DASH) outcome questionnaire (Beaton et al., 2001) and the less well-known assessments with functional items similar to those of the DASH, such as the Upper Extremity Functional Index (Stratford, Binkley, & Stratford, 2001) and Shoulder Pain and Disability Index (Bot et al., 2004). The third option is to assess function at the impairment level (i.e., loss or abnormality of anatomical structure or function; WHO, 2002) using measures such as the Minnesota Rate of Manipulation Test or Purdue Pegboard Test (Tiffin & Asher, 1948).

Although many assessments have been shown to be reliable and valid (Malouin, Pichard, Bonneau, Durand, & Corriveau, 1994; Poole & Whitney, 1988; Schulzer, Mak, & Calne, 2003; Tiffin & Asher, 1948; van der Lee, Beckerman, Lankhorst, & Bouter, 2001; Wolf et al., 2001) and granting agencies have spent millions of dollars to support their development (National Institutes of Health, 2005), they are not often used in the clinic, perhaps
because the assessments do little to inform day-to-day clinical practice (Woodbury & Velozo, 2005). Sometimes they are used simply to meet accrediting agencies’ requirements or reimbursement criteria (Woodbury & Velozo, 2005). Moreover, many clinicians view completing assessments as taking valuable time away from therapy (Woodbury & Velozo, 2005).

One reason that assessments fail to inform clinical practice is that they do not clarify the tasks patients can actually perform and the tasks that should be set as the next best goal for patients (Woodbury & Velozo, 2005). A total score is frequently calculated from patient responses; however, the total score is not directly interpretable because the actual tasks the patient is unable to complete are obscured. To be useful in informing clinical decision making (e.g., goal setting, treatment planning), an assessment must describe the abilities of a patient in a format that can be easily interpreted by the clinician.

To illustrate the way in which assessment total scores are unclear about actual ability, imagine a patient with a score of 70 (of a possible of 100) on the DASH outcome questionnaire, one of the most commonly used upper-extremity assessments (Woodbury & Velozo, 2005). This score of 70 tells the clinician little about the problems the patient is experiencing and what goals should be set to help conquer the challenges he or she faces. Different problems could lead to the same final score. For example, a score of 70 can be achieved by a patient with difficulties in fine motor tasks as well as by a patient with difficulties in gross motor tasks. Moreover, a total score does not tell the clinician the extent to which pain is a contributing factor or help set treatment goals for the client (Woodbury & Velozo, 2005).

A major disadvantage of using DASH total scores is that missing data become a problem. On the basis of the guidelines set for the DASH, at least 27 of the 30 disability/symptom questions must be completed for a score to be calculated (Solway, Beaton, McConnell, & Bombardier, 2002). When <30 questions are answered (e.g., questions that do not apply to the individual’s situation are left blank), however, missing data contribute to a lower total score. In this situation, a lower score does not indicate lower functional ability. Although missing data can be accounted for with the help of statistical programs, such statistical programs are not available in the clinic (Linacre, 1997).

A greater total score on the DASH indicates higher functional ability. It is not clear, however, how high a score must be to indicate that a person can perform a particular daily task. Information regarding ability to perform upper-extremity tasks in general can be provided by normative values. Normative values on DASH scores have been reported on 716 nonclinical working adults who were used as a reference group for a clinical population (Jester, Harth, & Germann, 2005). In that study, however, the DASH was divided into two parts (i.e., items representing impairment and items representing activity limitations), and scores were reported for each part instead of one total score for all DASH items being reported.

The only evidence available on association between total DASH score and functional ability was reported recently by our research group (Lehman, Sindhu, Shechtman, Romero, & Velozo, 2010). We found that a 20-point or more change in score on the DASH indicates a “true” change in function or symptoms. We used a sensitivity–specificity analysis (receiver operating characteristic [ROC] curves) to reach this conclusion. The ROC curve is a plot of false positives (the rate of negative cases incorrectly identified as positive by an assessment) along the x axis against true positives (the rate of positive cases correctly identified by an assessment) along the y axis (Shechtman, 2001). The optimal cutoff value of a ROC curve allows a decision to be made on the change score that best determines a change in clinical status considering the trade-off between sensitivity and specificity. Sensitivity is defined as the ability of an assessment to detect a clinical change when it actually exists, whereas specificity is defined as the ability of an assessment to detect the absence of a clinical change when the change actually does not exist (Portney & Watkins, 2000).

A possible solution to the problem of assessments that do little to inform clinical practice involves redesigning assessments using item response theory (IRT) methodologies. Rasch analysis, the one-parameter IRT model, can provide a clear picture of which items on an assessment will provide the right challenge for the patient (Kielhofner, Dobria, Forsyth, & Basu, 2005). That is, instantaneous evaluation of what the patient can and cannot do is made possible by placing his or her ability measure and item difficulty estimates on the same continuum (Linacre, 1997). Output from this analysis has been termed a keyform by Linacre (1997), who first proposed the idea. The keyform provides a visual display of expected response patterns for items on a given measure (Linacre, 1997).

Several concepts of the Rasch measurement model underlie the creation of keyforms. First, items on an assessment must be conceptualized as representing one idea or construct. If this assumption (i.e., that all items contribute to one construct) is upheld, Rasch analysis can be used to turn ordinal data into equal interval data with a unit of measure called a logit (log-odds unit). Therefore, the items contributing to the construct being measured
can be seen as making up a ruler with each item’s difficulty estimate (logit measure) falling at a specific marking. Person ability estimates (also calculated in logits on the basis of the patient’s responses to the items) can be placed on the same continuum, allowing for the direct comparison of patient ability with item difficulty (Kielhofner et al., 2005). These patient ability estimates indicate that someone possesses more or less of this construct. For instance, on the DASH, the upper-extremity ability will be less for a patient with a person measure of –2 than for a patient with a person measure of 2, regardless of whether both patients completed all the questions.

The purpose of this study was to demonstrate, using the DASH, how Rasch methodologies could be used to construct a clinically useful data collection form. Another purpose was to demonstrate how to use the data collection form to set short- and long-term goals as well as to show functional change from admission to discharge. Research questions were as follows: (1) Could Rasch methodologies be used to create a data collection form that would enhance clinical interpretation of the DASH? (2) Could this form be used to show functional change and to guide goal setting and treatment planning?

Method

Sample

Our sample of 960 participants was obtained from Focus on Therapeutic Outcomes, Inc. (FOTO; Knoxville, TN). FOTO gathers data using the Patient Inquiry® system, a computer software program developed by FOTO. Patients see questions on a computer screen and make a selection with the computer keyboard and mouse. FOTO takes multiple steps to ensure data integrity and reliability, including a fee schedule that promotes 100% patient participation and procedures that minimize content and selection bias (About FOTO, 2005). The selection criteria for participants were completion of the DASH at admission and discharge from outpatient therapy. All study participants completed all DASH questions (i.e., there were no missing DASH data). Participants were treated at orthopedic outpatient clinics around the United States; locations (urban vs. rural) and clinic ownership (physician vs. therapist) varied. The most commonly experienced impairments involved the shoulder (51%) or neck (20%). Symptom acuity was indicated by having the participants check chronic, subacute, or acute. Thus, interpretation of time periods associated with each category may have varied. More than half of the participants, however, reported that they suffered from chronic conditions. (A table summarizing demographic information for the sample is available online at www.ajot.aotapress.net; navigate to this article, and click on “supplemental materials.”)

Assessment

The DASH is a self-report inventory that measures the level of upper-extremity impairment and limitations in activities and participation. The assessment consists of a 30-item disability and limitation questionnaire and two optional modules: (1) a four-item sports and music questionnaire and (2) a four-item work questionnaire. The 30-item disability—limitation questionnaire was examined as part of this study. Items related to functional limitation typically ask the patient how much difficulty he or she has with a certain task. Responses range from 1 (no difficulty) to 5 (unable). Items related to symptoms ask respondents to rate their symptoms in the past week on a 5-point scale ranging from 1 (none) to 5 (extreme).

Scoring the DASH may be completed if no more than three items are unanswered. A score from 0 (no disability) to 100 (severe disability) is obtained by averaging the completed items, subtracting 1 from the average, and multiplying the difference by 25. For example, if a participant scores 5 on all the items, the average is 5, 5 – 1 = 4, and 4 × 25 = 100.

Preliminary Analyses—Exploratory Factor Analysis

In a preliminary analysis, an exploratory factor analysis (EFA) was conducted on the DASH. Because of the ordinal nature of the data, the EFA was conducted on the polychoric correlation matrix with weighted least square parameter estimates (Mplus software Version 4.21; Muthen & Muthen, 1998–2007). EFA identifies the number of factors (i.e., latent traits) that best explain the pattern of covariance underlying a set of measured variables (e.g., item responses; Brown, 2006). An oblique rotation method called promax was used because factors emerging from the data were expected to be correlated. Three factors had eigenvalues (λ) >1.0 (18.40, 1.56, and 1.54). All items had loadings of >0.40 on one factor, except for two of the items (sexual activities and transportation), which did not load on any of the constructs and thus were not included in the measures. The constructs identified from factor loadings included gross motor tasks, fine motor tasks, and symptoms. This study treats the three constructs as separate measures.

Generating Rasch Keyforms

We generated a keyform for each of the three constructs: gross motor tasks, fine motor tasks, and
symptoms. To make the keyform clinically meaningful, we used UMEAN and USCALE commands in the Winsteps Rasch software program (Linacre, 2005) to convert the raw person scores to person measures on a scale of 0–100. The UMEAN scores were calculated to be 48.991 for gross motor items, 47.301 for fine motor items, and 2.239 for symptoms items. The USCALE scores were calculated to be 7.949 for gross motor items, 9.003 for fine motor items, and 8.366 for symptoms items. In addition, because each item’s difficulty calibration varied from admission to discharge (i.e., as patients’ conditions changed, item difficulty estimates changed), alterations in functional ability were observed by making person measures at discharge comparable to those obtained at admission. Person measures at discharge were obtained by anchoring item average difficulties and average rating scale step estimates at admission values. That is, items were anchored at the difficulty of responding “moderate difficulty” to an item (item average difficulty). Simultaneously, items were anchored at the estimated value for the transition from one rating scale choice to the next, for example, from “mild difficulty” to “moderate difficulty” (average rating scale step estimate).

**Creation of Data Collection Forms**

The keyform output from Winsteps was used as the basis for creating the DASH data collection forms. The Winsteps form was modified by adding patient instructions at the top, moving items to the left of the response choices, and adding descriptions above the response choice numbers. Also, a percentage person ability scale ranging from 0–100 was placed at the bottom of the form. Moreover, the form was shaded to improve its readability.

To illustrate how a completed form looks, the gross motor data collection form was filled out according to responses of a randomly selected patient. Patient responses at both admission and discharge were circled on a form to show how patient progress might be observed using the form. On the admission form, possible goals were indicated by placing a box around potential short- and long-term goal activities. The discharge form illustrates whether these goals had been achieved.

**Results**

Three data collection forms were created for the DASH from the keyforms produced by Winsteps Rasch analysis program (Linacre, 2005). These forms are shown in Figures 1–3. Patient instructions are as follows: “Please rate your ability to do the following activities in the last week by circling the number below the appropriate response.” Patient instructions are presented at the top of each form. Items are located at the left in order of decreasing difficulty at admission. Rating scale choices (1–5) are presented to the right of each item with descriptions of what ratings mean above the number responses. The rating scale choices are placed at a location relative to the person ability scale at the bottom. For example, a patient with a person ability measure of 50 would be most likely to rate the item “Make a bed” as mild difficulty and rate “Carry a heavy object” as moderate difficulty (see Figure 1). A patient with a person ability measure of 30 would be most likely to rate both “Make a bed” and “Carry a heavy object” as severe difficulty. A patient with a person ability measure of 90 would most likely rate both items as no difficulty.

The gross motor data collection form of the DASH was filled out according to responses of a patient at admission and discharge. The form on the left has admission responses circled, whereas the one on the right has discharge responses circled (Figure 4). As can be seen in Figure 4, this patient was unable to complete all but three gross motor activities (washing and drying hair, pushing open a heavy door, and making a bed). The patient was able to perform the three tasks but with severe difficulty. Short-term goals for this person might include carrying a shopping bag or briefcase, washing his or her back, and increasing shoulder range of motion for tasks such as placing an object overhead and changing a lightbulb. Long-term goals might include increasing arm strength for tasks such as carrying a heavy object, gardening or doing yard work, heavy household chores, and recreational activities that require free arm movements and taking force or impact through the arm.

At discharge, this patient had met all but one short-term goal (placing object on shelf above the head) and had made progress toward all long-term goals. Using a functional approach to treatment (such as one used in constraint-induced movement therapy), this patient might be asked to perform activities such as reaching overhead to get objects out of a cabinet, carrying objects from one side of the room to another, or throwing a Frisbee.

**Discussion**

The purposes of this study were to create a clinically useful data collection form for the DASH, to illustrate how patient responses would look on the data collection form at admission and discharge, and to demonstrate how the forms can be used in setting patient goals. On the basis of the results of preliminary factor analysis, the DASH was divided into three subscales: (1) gross motor...
tasks, (2) fine motor tasks, and (3) symptoms. Although most items easily fit into one of the three groupings, two items (sexual activity and transportation) did not. Perhaps the items are poorly worded or lead to different interpretations.

The keyform output from the Winsteps Rasch analysis program (Linacre, 2005) was used as the basis for creating the DASH data collection forms. These data collection forms offer several advantages over the forms used in clinics today. First, this type of data collection form makes unexpected responses easy to spot (Kielhofner et al., 2005; Linacre, 1997). For example, if a patient has problems with all the items requiring shoulder range of motion (e.g., washing and blow drying hair, changing a lightbulb) but not on the more difficult items, this information can be diagnostic. In such cases, the therapist might assume the patient has a shoulder injury that is limiting shoulder range of motion.

Another advantage of this type of data collection form is that it can be used in conjunction with impairment and performance assessments to aid goal setting and treatment planning. The forms filled out with actual patient data illustrate this benefit. Examining such a form gives a sense of what activities and issues an individual completes with difficulty. For example, on the forms created and used to demonstrate goal setting at admission, it is clear that this patient has some range-of-motion issues (because of difficulty with the item “Wash your back”) and some problems with strength (because of indicated difficulty with the item “Carry a heavy object [>10 lbs]”). Using

---

**Figure 1. Gross motor data collection form.**

*Note.* Activity ratings (1–5) are placed above corresponding person ability measures. Therefore, for difficult items ratings are shifted farther to the right than for easier items.

**Figure 2. Fine motor data collection form.**

*Note.* Activity ratings (1–5) are placed above corresponding person ability measures. Therefore, for difficult items ratings are shifted farther to the right than for easier items.

---

The American Journal of Occupational Therapy

this information, the therapist can help the patient achieve improved range of motion and strength. On reevaluation, the clinician can once again see at a glance whether the items have improved and areas in which new goals might be set (i.e., where lower numbers are still being circled). New goals for this patient might include increasing ability to participate in more difficult recreational activities, such as golf or badminton. Finally, at discharge, treatment effectiveness can be determined by whether the patient is circling response choices to the right of the form (4s and 5s).

Several studies have discussed the creation of keyforms using Rasch methodologies. These studies have taken a slightly different focus from that of the current study. Earlier work by Linacre (1997) and Kielhofner et al. (2005) focused on the keyform as a way of obtaining “instantaneous measurement.” These authors focused their discussion on how using the keyform, without the use of computer software, a line can be drawn to represent an overall person measure for a patient and how patterns of responses can be used diagnostically. Linacre (1997) created a keyform based on data on the FIM™ (Keith, Granger, Hamilton, & Sherwin, 1987), whereas Kielhofner used data from the Occupational Performance History Interview–2nd Edition (Kielhofner & Henry, 1988; Kielhofner, Mallinson, Forsyth, & Lai, 2001). Another approach in using keyforms was demonstrated by Woodbury and colleagues (2007). Keyforms created from data on the Fugl-Meyer Assessment of the upper extremity were used to validate the consistency of the pattern of responses across study participants. Using the keyform, it was possible to verify that participants of differing abilities were scoring higher on easier items and lower on harder items. Evidence of this scoring pattern indicates that the Fugl-Meyer Assessment retains its structure when measuring participants across the ability range.

Work by Avery, Russell, Raina, Walter, and Rosenbaum (2003), using data collected on the Gross Motor Ability Estimator (Russell et al., 1989) demonstrated how a computer program could be used to generate output in a form similar to that of a keyform produced by Winsteps. The item hierarchy that they obtained shows a clear pattern of gross motor development. For instance, the difficulty hierarchy they found showed that in a baby, head lifting occurs before pulling to a sit, crawling follows attaining a sitting position, and walking and hopping come later in development. To obtain the keyform output described in the Avery et al. (2003) study, clinicians would be required to enter data collected on an assessment into a computer. The authors, however, reported that >50 therapists in Ontario said that they would be interested in using such a program (Avery et al., 2003). The advantage of a data collection form designed using the same

---

Figure 3. Symptoms data collection form.

*Note.* Activity ratings (1–5) are placed above corresponding person ability measures. Therefore, for difficult items ratings are shifted farther to the right than for easier items.
methodologies is that no computer entry would be required of the therapists.

This study represents the first time that a data collection form has been created for the DASH on the basis of keyforms generated through Rasch methodologies. Moreover, no published studies on keyforms present longitudinal data (i.e., data at both admission and discharge). Few studies have introduced the idea of this type of data collection form in occupational therapy literature and for use in occupational therapy clinic settings.

This study has several limitations. Although the reliability and validity of the original DASH have been well established (Atroshi, Gummesson, Andersson, Dahlgren, & Johansson, 2000; Beaton et al., 2001; De Smet, 2004; Hudak, Amadio, & Bombardier, 1996; Schuind et al., 2003), it is not known whether the psychometric properties of the DASH change when the items are reordered according to relative challenge and the rating scale modified so that higher numbers indicate more ability instead of more disability. Of concern is that

---

**Figure 4. Gross motor collection form with patient data: Illustrating goal setting.**

*Note.* Activity ratings (1–5) are placed above corresponding person ability measures. Therefore, for difficult items ratings are shifted farther to the right than for easier items.
the ordering of items with the most difficult at the top (read and answered first by the patient) and least difficult items at the bottom (read and answered last by the patient) might influence how the questions are perceived and answered. Such an influence of item ordering has been a well-known and well-studied phenomenon in the psychological literature for many years (e.g., Guertin, 1954; Stone, 1965). If patients subconsciously recognize that the activities in question are getting easier, they may start to respond accordingly without critically thinking about how challenging the task really is for them.

Also of concern is that on the keyform produced using Rasch measurement methodologies, the difficulty hierarchies differ slightly when they are produced using admission data from when they are produced using discharge data. If the difficulty ordering of items on an assessment were to vary widely from admission to discharge, this ordering could compromise the usefulness (and validity) of a data collection form produced using the methods in this study. Our preliminary analyses, however, showed that with a larger sample size, DASH item difficulties from admission to discharge were stable enough to not affect person ability estimates. In addition, in the current study, the concern of differing admission and discharge difficulty estimates was handled by obtaining person measures at discharge after anchoring on the basis of item and step difficulties at admission. The admission keyform was then used to produce the data collection form. This process was based on the thinking that therapists would want to see how much a patient’s ability changed from the point of admission. One further problem with this anchoring method was that because difficulties were anchored on the basis of admission data, many of the participants ended up with maximum person ability estimates at discharge.

In conclusion, a data collection form like the one presented in this article has many positive implications. First, as demonstrated by the completed forms in this study, this type of assessment form could be clinically useful to therapists, aiding in goal setting and treatment planning. Second, once such a form was adopted and widely used in clinics, it could be further used to justify reimbursement by insurance companies and by accreditation agencies. The clear illustration of improvement shown by this type of form would make an impressive statement about the possibilities of making functional gains from treatment. Finally, DASH data collection forms, like the ones presented in this study, would lead to possibilities for comparison of treatment effectiveness between therapists at different sites.

Future research with different populations and diagnoses is warranted. Because our sample consisted predominantly of patients with shoulder and neck impairments, one might wonder whether the item hierarchies would change if the sample were primarily made up of people with wrist and hand injuries. If the hierarchies were significantly different for those patients (as judged by differential item functioning and supplementary analyses), this finding would necessitate the creation of different data collection forms for each impairment group. In addition, studies investigating the influence of meaningfulness of daily tasks inquired about on the DASH would be insightful. The value that a person places on a task might affect his or her perception of difficulty of the task. This influence might show up on data collection forms, because the item responses might be unexpected (i.e., out of order according the normal hierarchy of item difficulty).

Moreover, future work comparing the psychometric properties of the created DASH data collection forms to already established reliability, validity, and responsiveness studies of the original DASH is necessary. Therapist perceptions of the usefulness of such data collection forms would be beneficial. Clinicians might suggest further modifications that would aid the treatment planning process. Actual trials in the clinic with these forms are needed to determine whether the data collection forms are useful in actual practice.

Acknowledgments
We thank Dennis Hart, Director of Consulting and Research, Focus on Therapeutic Outcomes, Inc., for use of FOTO’s database.

References


