Modification of the Upper Limb Functional Index to a Three-point Response Improves Clinimetric Properties

Charles Philip Gabel, MSc
Faculty of Science, Centre for Healthy Activities, Sport and Exercise, University of the Sunshine Coast, Queensland, Australia

Lori A. Michener, PhD, PT, ATC
Department of Physical Therapy, Virginia Commonwealth University, Virginia, USA

Markus Melloh, MD, MPH
Section of Medical and Surgical Sciences, Department of Orthopaedic Surgery, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

Brendan Burkett, PhD
Faculty of Science, Centre for Healthy Activities, Sport and Exercise, University of the Sunshine Coast, Queensland, Australia

ABSTRACT:
Study Design: Observational two-stage. Introduction: To achieve optimal clinimetric properties for outcome measures, both practical and psychometric, ongoing improvements are required.
Purpose of the Study: To evaluate if the Upper Limb Functional Index (ULFI) clinimetric properties are improved by modification to a three-point response option and to verify the factor structure.
Methods: Stage 1, calibration (n = 139) used ULFI dichotomous responses, and stage 2, validation (n = 117) used a three-point response option. The clinimetric properties were compared in physical therapy outpatients with the QuickDASH as the reference standard. Repeated measurements were made at two to four weekly intervals.
Results: The ULFI three-point response option improved reliability [intraclass correlation coefficient (2,1) = 0.98], internal consistency (α = 0.92), QuickDASH concurrent validity (r = 0.86), and responsiveness. Minimal detectable change (90% confidence interval) was 7.9%, and factor structure was unidimensional. Missing responses were 0.5%, and practical characteristics were unchanged.
Conclusions: The enhanced reliability and reduced errors with unchanged practicality demonstrate the ULFI improvements through modification to a three-point response option.
Level of Evidence: 2c.

Upper limb function and impairment are assessed by a variety of patient-reported outcomes (PROs) that reflect health at the activity and participation level.1–3 Recent systematic reviews found that no current PROs had positive ratings for all clinimetric properties,4 and further research was needed.5 The original Upper Limb Functional Index (ULFI) published in 2006 with a dichotomous response option6 was not included in these reviews because it was published after the review inclusion cut-off date.4 The original ULFI was concurrently validated with multiple response option PROs, the Disabilities of the Arm, Shoulder, and Hand (DASH)7 and the Upper Extremity Functional Index.8 Comparative analysis showed the original ULFI as the preferred PRO measure, mostly because of enhanced practical characteristics; however, the ULFI had slightly lower reliability, responsiveness, and higher error with change scores.6 These shortcomings were thought to be because of only “Yes” or “No” response options’ without an “intermediate” option.10,11 Without this third option, an individual’s response may become less precise and inconsistent.12 To improve the original ULFI psychometric properties and provide a three-point response option (ULFI3-pt), the questionnaire instructions required modification and the clinimetric properties of psychometrics and practicality reassessed concurrently with an accepted criterion standard.

The QuickDASH was selected as the criterion, as it was advocated for the assessment of upper limb musculoskeletal conditions.13–16 The QuickDASH
was derived from the DASH by the extraction of 11 items. This improved practicality and item redundancy, that had reduced the DASH clinimetric summary performance. Clinimetric properties are critical for a PRO to be accepted by patients, clinicians, and researchers. These include the psychometric (such as reliability, internal consistency, validity, change scores, responsiveness, and factor structure) and the practical (such as readability, administrative burden through completion and scoring, and missing responses) properties.

A questionnaire that is modified from a dichotomous to a three-point response option may provide a balance between practicality and improved clinimetric performance. This has been successfully demonstrated in research on dichotomous PROs.

**PURPOSE OF THE STUDY**

1. To determine whether psychometric and practical characteristics are improved when the original ULFI is modified from dichotomous to a three-point response option (ULFI3-pt).
2. To investigate the factor structure of the ULFI3-pt.

**METHODS**

Modification of the Original ULFI to a Three-point Response

To modify the wording of the original ULFI and produce the ULFI3-pt, two focus groups were formed. The first had five patients with different upper limb conditions; the second had five clinicians that included two physical therapists, two certified hand therapists, and one occupational therapist. Each group independently developed methods to provide the ULFI3-pt with the desired three-point response option where the third point was central between “Yes” and “No.” The consensus decision was to add the statement, “If an item partly describes you, Use a Half (1/2) Mark.” No other changes were made (Figure 1).

Next, to assess this new format, a group of patients (n = 20) consisting of four participants from five separate clinics each completed four questionnaires: the ULFI3-pt (with a one-box response option); the QuickDASH; the ULFI3-pt with a three-box response option; and an 11-point (0–10) numeric rating scale (NRS) practicality questionnaire. This practicality questionnaire was anchored at 0 (“Not at all”) to 10 (“Yes”) and contained four questions: “Rank your assessment of each of the three questionnaires. Was it …” 1) “… difficult to complete,” 2) “… confusing,” 3) “… requiring further explanation,” and 4) “… appropriate to [their] condition.”

**PRO Questionnaires**

The ULFI3-pt is a single page, 25-item upper limb regional PRO, with the response options of “Yes”/“Half”/“No” and scored by assigning 1 point for each “Yes,” 0.5 points for each “Half,” and 0 points for each “No.” The total points are added and multiplied by four for a total score of functional limitation, 0 (no limitation) to 100 (maximum limitation).

The QuickDASH is a two-page, 11-item, shortened version of the original 30-item DASH. The response options are a 1–5 Likert scale with a raw score range of 11–55. The raw score is converted to a percentage, 0 (no disability) to 100 (most severe disability) and allows for one missing response. The QuickDASH has an additional third page, containing two optional modules for “work” and “sports/performing arts.”

**Setting and Participants**

Participants (described in Table 1) with upper limb musculoskeletal conditions were consecutively recruited from physical therapy outpatient clinics in
Australia and the United States. Inclusion criteria were any upper limb condition that included soft-tissue injury, postsurgery, lymphedema, fractures, chronic regional pain, and trauma. Exclusion criteria were 18 years of age, difficulty with English language comprehension, and cognitive impairment. Participants completed the PRO questionnaires at initial evaluation, those receiving ongoing treatment were measured

**UPPER LIMB FUNCTIONAL INDEX**

**DATE:**

**NAME:**

**INJURY:**

☐ LEFT ARM ☐ RIGHT ARM

**PLEASE COMPLETE:** Your arm may make it difficult to do some things you normally do. This list contains sentences people use to describe themselves with such problems. Think of yourself over the last few days. **If an item describes you mark the box. If not leave the box blank. If an item partly describes you Use a Half (½) Mark.**

**DUE TO MY ARM:**

| 1. I stay at home most of the time. |
| 2. I change position frequently for comfort. |
| 3. I avoid heavy jobs e.g. cleaning, lifting more than 5kg or 10lbs, gardening etc. |
| 4. I rest more often. |
| 5. I get others to do things for me. |
| 6. I have the pain / problem almost all the time. |
| 7. I have difficulty lifting and carrying (e.g. bags, shopping up to 5kg or 10lbs). |
| 8. My appetite is now different. |
| 9. My walking or normal recreation or sporting activity is affected. |
| 10. I have difficulty with normal home or family duties and chores. |
| 11. I sleep less well. |
| 12. I need assistance with personal care e.g. washing and hygiene. |
| 13. My regular daily activities (work, social contact) are affected. |
| 14. I am more irritable and / or bad tempered. |
| 15. I feel weaker and / or stiffer. |
| 16. My transport independence is affected (driving, public transport). |
| 17. I have difficulty putting my arm into a shirt sleeves or need assistance dressing. |
| 18. I have difficulty writing or using a key board and / or ‘mouse’. |
| 19. I am unable to do things at or above shoulder height. |
| 20. I have difficulty eating and / or using utensils (e.g. knife, fork, spoon, chop sticks). |
| 21. I have difficulty holding and moving dense objects (e.g. mugs, jars, cans). |
| 22. I tend to drop things and / or have minor accidents more frequently. |
| 23. I use the other arm more often. |
| 24. I have difficulty with buttons, keys, coins, taps / faucets, containers or screw-top lids. |
| 25. I have difficulty opening, holding, pushing or pressing (e.g. triggers, lever, heavy doors). |

**ULFI SCORE:** To score the upper part - add the marked boxes:

| TOTAL (ULFI points) | 100 Scale (x 4) 100 – Total = % |

**MDC (90% confidence):** 7.9 % or 1.9 ULFI points. Change less than this may be due to error.

**FIGURE 1. Upper Limb Functional Index.**

Australia and the United States. Inclusion criteria were any upper limb condition that included soft-tissue injury, postsurgery, lymphedema, fractures, chronic regional pain, and trauma. Exclusion criteria were <18 years of age, difficulty with English language comprehension, and cognitive impairment. Participants completed the PRO questionnaires at initial evaluation, those receiving ongoing treatment were measured...
again at two weekly intervals for six weeks, then four weekly thereafter until discharge. Status was classified as acute—injured within the previous six weeks; subacute—six to 12 weeks; and chronic—greater than 12 weeks. This study was approved by the University of the Sunshine Coast Human Research Ethics Committee, and all the participants completed an informed consent.

Stage 1, Calibration

A retrospective analysis was performed of data from the original study of the ULFI6 to assess concurrent validity between the original ULFI and the QuickDASH criterion (as 11 items extracted from the DASH7,18,27). To perform this analysis, we extracted the 11 items from the DASH to create QuickDASH scores. Participants (n = 139) were recruited from nine physical therapy outpatient centers in three different Australian states, and completed the ULFI and DASH at time points as described above. Demographic details are presented in Table 1.

Stage 2, Validation

A prospective investigation of participants (n = 117) recruited from seven private physical therapy outpatient clinics, six Australian and one American (Table 1). All the participants completed the ULFI3-pt, and their responses used to assess internal consistency, ceiling and floor effects, missing responses, and to assess factor analysis structure of

### TABLE 1. Participant Demographics for Upper Limb Functional Index: Stage 1, Calibration and Stage 2, Validation

<table>
<thead>
<tr>
<th>Demographic Data</th>
<th>Stage 1 (Gabel et al., 2006)</th>
<th>Stage 2 (Present Study)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants (n)</td>
<td>139</td>
<td>117</td>
</tr>
<tr>
<td>Responses (n)</td>
<td>211</td>
<td>366</td>
</tr>
<tr>
<td>Age (yr)</td>
<td>48.4 ± 15.6</td>
<td>49.9 ± 16.1</td>
</tr>
<tr>
<td>Gender: % female</td>
<td>54</td>
<td>35</td>
</tr>
<tr>
<td>Dominance: % right</td>
<td>77</td>
<td>97</td>
</tr>
<tr>
<td>Injury Duration (wk)</td>
<td>24.5 ± 28.8</td>
<td>13.4 ± 17.3</td>
</tr>
<tr>
<td>Time range (wk)</td>
<td>1–433</td>
<td>1–80</td>
</tr>
<tr>
<td>Work status (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>61</td>
<td>64</td>
</tr>
<tr>
<td>Retired</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>Unemployed</td>
<td>39</td>
<td>4</td>
</tr>
<tr>
<td>Injured at work (%)</td>
<td>40</td>
<td>28</td>
</tr>
<tr>
<td>On workcover (%)</td>
<td>30</td>
<td>26</td>
</tr>
</tbody>
</table>

FIGURE 2. Flow chart of calibration from reanalyzed data (stage 1) and prospective validation (stage 2).
the ULFI3-pt. Subgroups formed assessed reliability (n = 47), responsiveness (n = 95), and concurrent validity with the QuickDASH (n = 67) by consecutive allocation of participants within each clinic, and the total participants exceeded the minimum required number. Subregions were distal, n = 14; central, n = 21; proximal, n = 65; and general, n = 17.

**Data Analysis—Psychometric Characteristics**

**Distribution and Normality**

This was determined from the baseline scores for each PRO from histogram inspection and the one-sample Kolmogorov–Smirnov test.28

**Internal Consistency**

Cronbach’s alpha coefficient (α) was used with an optimal range of 0.90–0.9425,30 for baseline measurements.

**Reliability**

Measures taken at baseline and 72 hours later during a period of non-treatment6 were compared using the Type 2,1 intraclass correlation coefficients (ICCs).31 An 11-point NRS “global rating of change” was completed as a reference criterion6,52 to determine those participants who were “stable” and appropriate to use for reliability analysis. Participants (n = 47) who were “unchanged,” defined as a “Change = 0 ± 1” as the bounds of acceptance were included for the reliability analysis.6,7

**Responsiveness**

Participants (n = 95) who had repeated measures of the ULFI3-pt were used to assess responsiveness. Responsiveness was further classified by the time since injury, two weeks for acute participants, four weeks for subacute participants, and six weeks for chronic participants.6,7 The score change was measured with the indices of effect size (ES)33 and standard response mean (SRM).34,35 Some participants (n = 22) received no follow-up or were discharged before the stipulated repeated measurement time from which the last PRO score was recorded, and in these cases an “intention-to-treat” analysis was used.36

**Measurement Error**

The minimal detectable change (MDC) was calculated at the 90% confidence interval (MDC90) by initially calculating the standard error of the measurement (SEM) using the formula: $SEM = SD_{av} \sqrt{(1 – ICC)}$, where $SD_{av}$ was the average standard deviation (SD) of scores for all baseline and follow-up measures and ICC was the test–retest reliability coefficient.37 The MDC or the error associated with repeated measurements was determined using the formula: $SEM_{repeat} = \sqrt{2 \times SEM}$, which accounts for the error associated with both the initial and repeated measurements.38 The MDC90 is subsequently determined from multiplying by the Z value of 1.64, which corresponds to the 90% confidence bounds. The minimal clinically important difference (MCID) error value was calculated using a distribution-based method, with a minimum level of 20% change on the corresponding NRS.39–41

**Validity**

Face and content validity for the ULFI3-pt and QuickDASH were assessed through patient and clinician feedback from the practicality subgroup (n = 20) via the readability scores.22 Criterion validity was assessed using a Pearson coefficient from concurrent comparison of the ULFI3-pt with the QuickDASH total scores for the responses (n = 184) from those participants who completed both PROs. Construct validity was determined through longitudinal and discriminant validity. Two external criteria were used, self-rated change of health status in the affected arm of ≥2.0 points change on the 11-point NRS42–44 and a 12.5% change45,46 on the Patient-Specific Index.6 Both criteria were required to categorize a subject as improved or deteriorated.37 These changes in cut-off values ensured the MCID on the criteria standards.47 Discriminant validity used three criteria: 1) a statistical difference in the mean change scores between the two responsiveness test groups assessed with the paired t-test; 2) comparative analysis of those patients impaired (>20% and 2 × MDC90 of the DASH) and those “able to do everything they need to,” that is, discharged or recovered to a level less than the MDC17; 3) the presence of a higher “distal” subregion mean, a within region test.7,23

**Factor Analysis**

The baseline data for the ULFI3-pt and QuickDASH were assessed using maximum likelihood extraction, which required the assumptions of normality as opposed to the default method of principal component analysis that has no distributional assumptions.48 The loading coefficient absolute value suppression was set at 0.30.30,52 Factor extraction was determined by three methods: the scree plot curve “point of inflection”49, an eigenvalue cut-off of 1.050; and that ≥10% of variance was accounted for where average communality (after extraction) was ≥0.6.28,51 With determination of the number of extracted factors, the data were reanalyzed and verified by the forced solution method. Varimax rotation was used to demonstrate factor item loading where two or more factors were determined.28,51 A unidimensional
structure (the presence of a single underlying construct or theme) was required for a summated score to be valid.52

Data Analysis—Practical Characteristics, Readability, and Summary Performance

Nine essential areas of practicality were considered6,22,23,53 the initial five are self-evident: 1) being self-administered, 2) applicable across a variety of conditions, 3) related to a variety of disease severity levels, 4) relevant to defined populations, and 5) maximum length of one page. To determine the remaining four areas: the results from the practicality subgroup were used for 6) completion time, 7) scoring time, and 8) ease of understanding and completion; for 9) missing responses, the percentage was calculated from the total responses. Readability was assessed to quantify the ease of understanding and was ascertained from the Flesch–Kincaid scale, which assigns a score on the basis of the minimal grade level required to read and understand English text (range, 0–12) and should be ≥grade 7 for self-report questionnaires.54 It was calculated automatically from the word processor software grammar function and demonstrated as reliable and valid.55 Finally, summary performance was assessed using two scales: 1) the “measurement of outcome measures,” which dichotomously evaluated 25 essential clinimetric properties divided into four categories of methodological, practical, distribution, and general and scored on a 100% scale6 and 2) the “Bot” clinimetric scale considered 12 items under four response options of good, poor, doubtful, and unavailable, with the score dichotomously summed.22 Some of the cut-off classifications of the “Bot” scale were considered too conservative in this study: “Time to administer” was reduced from 10 to 3 minutes for completion and scoring, and “Readability and comprehension” were quantifiably defined as the seventh grade level; “Bot” used a subjective self-reported assessment. The Statistical Package for Social Sciences version 14.0 (SPSS Inc., Chicago, IL) was used for all analyses.

Sample Size

The required power for reliability and responsiveness were, respectively, 22 and 53 participants to provide an 80% confidence level in determining actual change56 of 10.5%. This value is greater than the established MDC90 for the original ULFI6 and the DASH,7,57 as there are no published values for the QuickDASH. For criterion investigation, 175 was determined as calculated using Meng’s test of significance and solving for n.8,58 For the factor analysis, >100 was required based on the assumptions of normality, consecutive sampling, and the a priori requirements for factor extraction.20,59

RESULTS

Psychometric Characteristics

These are presented for each PRO in both stages in Table 2 with the construct validity in Table 3. The values for the QuickDASH are invalid for summation to a single repeated score because of the bidimensional factor structure (the presence of two underlying constructs or themes).28,51 They are provided for comparison to the other PROs and previous QuickDASH studies.

Distribution and Normality

Normality was demonstrated through the Kolmogorov–Smirnov test (0.78, sig 0.80) and histogram inspection. The distribution of the baseline data for both the ULFI3-pt and QuickDASH covered the full range from “unaffected” 0% to “maximum impairment” 100% in both the stages. Only one

| TABLE 2. Methodological Characteristics of ULFI and QuickDASH Criterion |
|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Stage             | Reliability      | Internal         | Error Score     | Responsiveness  | Missing          |                |                |
|                   | Rxx              | Consistency Alpha| SEM (%)         | MDC90 (%)       | SD100 (%)        | ES              | SRM             | Percentage      |
| Development (2006) | 0.96             | 0.89             | 4.50            | 10.50           | 21.61            | 1.28            | 1.87            | <0.5           |
| Original ULFI     |                  |                  |                 |                 |                 |                 |                 |                |
| QuickDASH (extracted items) | 0.94 | 0.92 | 4.98 | 11.58 | 20.71 | 1.21 | 1.75 | 12.5 |
| Present study     |                  |                  |                 |                 |                 |                 |                 |                |
| ULFI3-pt          | 0.98             | 0.92             | 3.41            | 7.93            | 24.16            | 0.93            | 1.33            | <0.5           |
| QuickDASH         | 0.91             | 0.92             | 6.73            | 15.66           | 23.20            | 1.05            | 1.25            | 26.6           |

Rxx = Test–retest reliability coefficient; alpha = Cronbach’s alpha; ULFI = Upper Limb Functional Index; ICC = intraclass correlation coefficient; SEM = standard error of the measurement; MDC90 = minimal detectable change (90% confidence interval); SD100 = standard deviation at baseline (100% scale); ES = effect size; SRM = standard response mean. The QuickDASH psychometric properties are invalid. They are provided as a reference only.
patient reported a 100% level and this was on both PROs. The “Half Mark” response option was used in 69% of the ULFI3-pt responses by 83% of the participants. The validation of the ULFI 3-pt baseline responses showed a more evenly distributed score range and slightly improved histogram shape compared with the calibration stage with the dichotomous response.

Validity

Criterion validity was consistent in both the stages; calibration, $r = 0.85$ and validation, $r = 0.84$. Discriminant validity was demonstrated for the original ULFI and ULFI3-pt by the difference in mean scores between baseline and repeated measures in the responsiveness group with a significant t-statistic (Table 3); that recovered patients scores were below the MCID of those not recovered; there was a difference in mean scores for subregions where a higher “distal” mean was demonstrated, for the ULFI3-pt that was not found for the QuickDASH (Table 4).

Factor Structure

The ULFI3-pt was shown to be suitable for factor analysis, as the correlation matrix had a Kaiser–Meyer–Oklin value of 0.912 and a significant Barlett Test of Sphericity ($p < 0.001$). The three a priori criteria were met and indicated a unidimensional structure. The scree plot inflection point was found at the first value and only one eigenvalue $>1.0$ accounted for variance $>10\%$ (total $= 33.4\%$). It was noted that seven factors had eigen values $>1.0$ and accounted for 64.9% of variance. Four factors were between 0.5 and 1.0, and the remainders were below 0.5%. The QuickDASH had a bidimensional structure.

Practical Characteristics

Completion and Scoring Times

For the ULFI3-pt, the completion time was $117 \pm 47$ seconds, and the QuickDASH was $95 \pm 33$ seconds. For the therapist scoring times, the ULFI3-pt required $16 \pm 4$ seconds and the QuickDASH $60 \pm 31$ seconds; however, in the presence of one missing response, the QuickDASH scoring time increased to $124 \pm 7$ seconds. Combined, the ULFI3-pt was $133 \pm 51$ seconds, and the QuickDASH ranged from $155 \pm 64$ to $219 \pm 40$ seconds.

Readability

The original ULFI and ULFI3-pt both had readability below the seventh grade level, the QuickDASH was grade 12. For the practicality questionnaire, by the 20 participant practicality group, the ULFI3-pt mean was lower than the QuickDASH for all the four questions but the differences were not statistically significant.

Missing Responses

These were minimal for the ULFI3-pt with two in the total pool of 366 ($<0.5\%$). This was consistent with the calibration group also at $<0.5\%$.6 The

---

**TABLE 3. Construct Validity Comparing Baseline and Repeated Scores for the Upper Limb Functional Index (ULFI) and QuickDASH**

<table>
<thead>
<tr>
<th>Tool</th>
<th>Sample Size (n=)</th>
<th>Baseline Mean</th>
<th>Repeat Test* Mean</th>
<th>Paired t-Statistic†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development (2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original ULFI</td>
<td>31</td>
<td>58.1 ± 23.0</td>
<td>41.3 ± 26.6</td>
<td>5.6</td>
</tr>
<tr>
<td>QuickDASH (extracted items)</td>
<td>29</td>
<td>58.2 ± 20.8</td>
<td>40.6 ± 23.1</td>
<td>5.7</td>
</tr>
<tr>
<td>Present study</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ULFI3-pt (pooled groups)</td>
<td>95</td>
<td>45.4 ± 24.2</td>
<td>26.3 ± 22.3</td>
<td>8.9</td>
</tr>
<tr>
<td>QuickDASH</td>
<td>64</td>
<td>44.4 ± 23.2</td>
<td>20.0 ± 16.1</td>
<td>4.2</td>
</tr>
</tbody>
</table>

*Repeated measures were made after a period of known natural healing: acute after two weeks; subacute after four weeks; and chronic after 13 weeks.
†p-value <0.0001 for all t-statistic measures.

**TABLE 4. Mean Scores by Upper Limb Subregion for Upper Limb Functional Index (ULFI) and QuickDASH**

<table>
<thead>
<tr>
<th>Subregion</th>
<th>Stage 1</th>
<th>Stage 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original ULFI</td>
<td>QuickDASH</td>
</tr>
<tr>
<td>Proximal—shoulder and upper arm</td>
<td>36.6 ± 22.6</td>
<td>46.7 ± 20.4</td>
</tr>
<tr>
<td>Central—forearm and elbow</td>
<td>36.6 ± 22.6</td>
<td>44.8 ± 18.7</td>
</tr>
<tr>
<td>Distal—hand and wrist</td>
<td>30.0 ± 29.7*</td>
<td>43.7 ± 27.0</td>
</tr>
<tr>
<td>General—whole arm</td>
<td>48.3 ± 22.0</td>
<td>43.9 ± 21.4</td>
</tr>
<tr>
<td>All—average for all the data</td>
<td>45.3 ± 23.6</td>
<td>45.3 ± 21.6</td>
</tr>
</tbody>
</table>

*Denotes higher distal mean as recommended.7,23
QuickDASH had 26.6% missing, of which 16.6% were from question ten.

Summary Performance

The original ULFI and ULFI3-pt were identical and scored, respectively, higher than the QuickDASH on both the “measurement of outcome measures” scale (ULFI = 96%, QuickDASH = 44%) and the “Bot” scale (ULFI = 12/12 or 100%, QuickDASH = 3/12 or 25%).

Clinic scores were not statistically different between the American and Australian participants, which enabled pooled analysis.

DISCUSSION

The ULFI3-pt demonstrated validity and reliability as a three-point response scale questionnaire. The improved reliability over the original ULFI reduced the measurement error, which in turn made the questionnaire more sensitive to change. The additional “Half Mark” response option was accepted by most of the participants in most of their responses. The similar demographic factors, psychometric values, and criterion validity between the calibration and validation stages indicated consistency between samples, that comparison was acceptable and that the QuickDASH was an appropriate criterion. This was supported by the measured responsiveness and reliability for the QuickDASH, which compared favorably to previous research findings. The ULFI3-pt clinimetric properties were demonstrated as preferable to the original ULFI, and the factor structure was determined as unidimensional.

Minor alterations to PROs are often made by researchers to improve the psychometric properties, reduce patient burden, and improve scale practicality. However, before a new version of any PRO can be adopted it must be validated in an independent investigation. Although the psychometric changes to the ULFI3-pt were minor, they have potentially far reaching consequences. The three-point response option provided interval rather than ordinal data through two critical perspectives: psychologically, the three required positions of “Yes,” “No,” and “Intermediate” were available, and statistically, the question response options were equally spaced individual interpretations that provided a summed score with normalized distribution. Improvements in reliability and responsiveness will enable clinicians and researchers to more efficiently determine if their intervention strategy was effective or not. This may save time, improve results, and comply with evidence-based medicine (EBM) standards.

The normalized distribution of the ULFI3-pt total and subregion scores was demonstrated. The maximum and minimum scores indicated no tendency to floor or ceiling effect. Consequently, the ULFI3-pt item constructs had adequate range to discriminate change. This discriminative capacity was demonstrated in two ways: differences between scores at different periods and through the higher “distal” mean. This latter capacity was critical as hand patients are recognized as having the highest impairment.

The ULFI3-pt internal consistency was marginally higher in the validation stage but remained below the 0.95 cut-off for item redundancy. The improved test–retest reliability led to improved sensitivity, lowered the SEM, and improved the MDC90 from 10.5% to 7.9%. Consequently, clinical and research application using the ULFI3-pt will have greater sensitivity for detecting change with interventions that may otherwise not show a valid effect. This may potentially reduce the required time to conclusively demonstrate change that has occurred and that an intervention was effective or not effective. The determined effectiveness or lack of effectiveness of intervention strategies is the foundation of EBM. However, it relies on the ability of PRO measures to determine when clinically meaningful changes have occurred in a patient’s status. The modification of the original ULFI to produce the ULFI3-pt resulted in a more sensitive three-point response option ULFI3-pt.

Responsiveness was lower in the validation stage, which may be because of its more diversely impaired sample. This was supported by the higher SDs found at baseline, on repeated measures and in the subsequent change scores. The higher levels of change were adopted to ensure that the MCID was achieved. This was an observational study and other possibilities that influence responsiveness and make change harder to detect could include different interventions by the treating therapists, the duration of follow-up (as an instrument is less responsive over shorter follow-ups), and the level of severity at baseline (as the amount of change varies between acute and chronic patients). Although responsiveness was lower in the validation stage, it reinforced construct validity, as both ES and SRM remained at a “high” magnitude of change, >0.80 level.

The factor analysis showed a unidimensional model for the ULFI3-pt with consistent levels in variance, which indicated the items formed one construct, upper extremity impairment. The QuickDASH bidimensional structure indicated that it had two underlying constructs. Despite a single dominant factor, the ULFI3-pt had six additional factors with eigenvalues >1.0 that accounted for a substantial percentage of variance, and 14 items scored below 0.50. This suggested that the ULFI3-pt may be shortened which would further reduce respondent burden and improve practicality. This would need to be determined by further research.
The maximum 100% score for both ULFI versions on the “Bot” scale\textsuperscript{26} and 96% on the “measurement of outcome measures” scale\textsuperscript{6} supported the previous finding of the original ULFI as preferred to the DASH. In particular, the completion and scoring times were highly efficient and missing responses were insignificant. Clinimetric assessment scales provided a means to compare summary performance between PROs that measure the same body region. These scales measured the presence of the components, not the actual values, and whether they were at an acceptable level. For example, reliability was high for both ULFI versions and scored the maximum on both summary performance scales, but the ULFI\textsubscript{3-pt} at 0.98 was improved compared with the original ULFI at 0.96. By contrast, for readability, the ULFI\textsubscript{3-pt} was much easier to read and comprehend with a seventh grade level compared with the 12th grade of the QuickDASH.

The QuickDASH was used as a criterion reference standard, and was shown to be comparable with the extracted DASH items from the calibration stage. The finding of a bidimensional structure for the QuickDASH indicates that a single summated score may be inappropriate.\textsuperscript{64} The bidimensional structure of two underlying constructs indicates that two sub-scores are appropriate, which can be then summed for a total score. The QuickDASH also demonstrated slightly lower reliability, responsiveness, and consequently sensitivity.

The slight variation in demographic data between the validation and calibration populations may be responsible for some differences in results. The calibration group compared with the validation group had higher female and unemployed representation but lower retired participants possibly because of the classification not being interpreted correctly. Geographic and economic drivers of the population regions may also have contributed. The calibration stage population involved predominantly mining, agriculture, military, and fishery workers with no retired participants; the validation stage population had mostly tourism and hospitality workers.

This research fulfilled the recommendations of the two most recent systematic reviews on upper limb PROs,\textsuperscript{2} as it provided further research on clinimetric properties and positive ratings for the ULFI\textsubscript{3-pt}. The study demonstrated that the ULFI\textsubscript{3-pt} improved sensitivity and reduced clinician burden. This study had limitations. Only patients from physical therapy outpatients were used, and there was no investigation of specific conditions, groups, or settings. The results cannot be generalized to other patients or settings.

Implications for Further Research

The consistency in the criterion validity between the ULFI\textsubscript{3-pt} and QuickDASH (which was validated in different condition specific populations), implied generalizability, but further validation is required. New research would be required that used repeated measures on equivalent and new population groups. The MCID should be determined independently from a statistically recommended method that used specific external criteria based on the patient’s symptoms and evaluated treatment interventions.\textsuperscript{40} With the potential to shorten the ULFI\textsubscript{3-pt}, perhaps to ten items, the demand on respondents and clinicians would be further reduced.\textsuperscript{65}

CONCLUSIONS

The ULFI\textsubscript{3-pt} improved the original ULFI psychometric properties without the loss of clinical utility and demonstrated a unidimensional factor structure. Practical characteristics were retained and a high overall performance score for both the “Measurement of outcome measures” and “Bot” clinimetric summary performance scales. These characteristics were both preferable and superior to the QuickDASH, which had questionable validity because of its bidimensional structure. The findings indicated that the ULFI\textsubscript{3-pt} is viable as a PRO measure for the determination of upper limb status and impairment in both the clinical and research settings.

Acknowledgments

This research was supported by the University of the Sunshine Coast. We thank all participating patients, general practitioners, and therapists for their time and effort. The study was approved by the Human Research Ethics Committee of the University of the Sunshine Coast. We wish to thank Lukas Staub for statistical editing and Selena Horner of Red Cedar Physical Therapy for her invaluable contribution of patient’s data.

REFERENCES


19. Institute for Work and Health. Development and testing of the


Record your answers on the Return Answer Form found on the tear-out coupon at the back of this issue. There is only one best answer for each question.

#1. The ULFI
   a. is being introduced as a new outcome measure in this issue of the JHT
   b. has been previously described in an earlier issue of the JHT
   c. was developed at the Fulbright Institute of the University of Louisiana
   d. is a research tool, not intended to be used clinically

#2. The 3-point response option proved to be
   a. more time consuming to administer
   b. less time consuming to administer
   c. less reliable with more errors
   d. more reliable with fewer errors

#3. The outcome measures were tested on a patient sample of
   a. 100
   b. 10
   c. 20
   d. 40

#4. Concurrent validity of the ULFI was determined by comparing it to the
   a. DASH
   b. QuickDASH
   c. SF36
   d. UEFI

#5. The author uses the acronym PRO to stand for
   a. patient rated outcome (questionnaire)
   b. patient response observation (questionnaire)
   c. perceived rank outcome (questionnaire)
   d. practical ranking of outcomes (questionnaire)

When submitting to the HTCC for re-certification, please batch your JHT RFC certificates in groups of 3 or more to get full credit.