ORIGINAL REPORT

EVALUATION OF THE PSYCHOMETRIC PROPERTIES OF THE UPPER LIMB SUBSCALES OF THE MOTOR ASSESSMENT SCALE USING A RASCH ANALYSIS MODEL

Kimberly J. Miller, PhD[1], Anita L. Slade, PhD[2], Julie F. Pallant, PhD[3] and Mary P. Galea, PhD[1]

From the [1]Rehabilitation Sciences Research Centre, Physiotherapy, Melbourne School of Health Sciences, University of Melbourne, Melbourne, Victoria, Australia, [2]Neurological Outcome Measurement Unit, Peninsula College of Medicine and Dentistry, University of Plymouth, Plymouth, UK and [3]School of Rural Health, University of Melbourne, Shepparton, Victoria, Australia

Objectives: To apply Rasch analysis to evaluate the psychometric properties of the composite score of the 3 upper limb subscales of the Motor Assessment Scale (UL-MAS) when administered in the acute/subacute phase post-stroke.
Design: Prospective data collection of UL-MAS scores.
Participants: Eighty individuals a mean of 64.8 days (standard deviation 53.3; range 4–193 days) following the onset of unilateral stroke.
Methods: All UL-MAS test items were administered in 30 participants assessed longitudinally over 3 occasions, and in 50 participants assessed on a single occasion. These 140 observations were pooled to be evaluated using Rasch analysis.
Results: With the elimination of the wrist radial deviation test item, the UL-MAS demonstrated uni-dimensionality with no significant test item response bias. The test item difficulty hierarchy was validated in the Upper Arm and Hand Movements subscales, but not in the Advanced Hand Activities subscale. The acceptable floor (14%) and ceiling (9%) effects and the high Person Separation Reliability Index (0.96) indicated that the scale was appropriately targeted to discriminate statistically between groups of acute/subacute stroke participants with differing upper limb motor recovery.
Conclusion: The findings support the psychometric properties of the composite UL-MAS score in this clinical population.

Key words: stroke; upper extremity; disability evaluation; validation studies.

INTRODUCTION

Up to 85% of individuals hospitalized following stroke will initially present with upper limb (UL) dysfunction (1) and many of these individuals never fully regain functional use of their arm and hand (2). An important step in addressing deficits and establishing effective UL rehabilitation programmes is the refinement of outcome measures so that they are valid and reliable in assessing the nature and extent of the problem and documenting changes in ability.

The 3 upper limb subscales of the Motor Assessment Scale (UL-MAS) were developed as clinical and research tools (3) objectively and reliably to quantify arm and hand motor recovery following stroke (4, 5). An ordinal score is assigned based on the performance of test items presumed to be ordered hierarchically by difficulty in each subscale; a higher score reflecting better UL motor function.

Ordinal-scored scales are commonly used in rehabilitation outcome measures; however, they are problematic when applied to investigate treatment efficacy, as parametric statistics cannot be used to compare change scores between experimental groups (6). A change score of 1, improving from a score of 2–3 cannot be assumed to be equivalent to improvement from a score of 5–6 on the same subscale. When ordinal scales, such as the UL-MAS, are administered to evaluate change associated with recovery or therapeutic interventions, psychometric properties additional to the reliability attributes of the scale, are important to establish. The scale should have construct validity and be appropriately calibrated for the clinical group in which it is intended to be administered. The test items should reflect the trait(s) of interest (in this case UL motor recovery), unbiased by the patients' personal attributes (for example their gender or age), and employ test items of appropriate difficulty to be sensitive and responsive to the abilities of the patients (7, 8).

Rasch analysis is a relatively new mathematical approach that can be applied to investigate the psychometric properties of ordinal scales (7). Rasch analysis uses a probabilistic model, which assumes that the easier the test item, the more likely it will be passed (or performed) by participants (or patients); therefore the more able the person, the more likely they will pass a test item compared with a less able person (9). Ordinal scales that fit the Rasch model expectations can be transformed into equal interval measurement, thus enabling parametric statistical analyses to be used (6). Assumptions regarding construct validity can be explored, to establish if the scale is measuring a single uni-dimensional trait (UL motor function) or is influenced by other constructs (e.g. range of motion, spasticity). Estimates for participant (or patient) abilities and the difficulty of the test items can be calibrated

on a common interval scale, a log odds ratio (logits) scale (10). The targeting of test item difficulty in relation to patient abilities can be examined, exposing ceiling and floor effects, as well as the capacity of the test items to distinguish between groups of patients with differing levels of ability (7). Response bias (referred to as Differential Item Function: DIF) related to the personal attributes of different subgroups of patients (e.g. gender, age) can be established and the difficulty hierarchy of test items in the scale can be evaluated (7).

Aamodt et al. (11) applied Rasch analysis to evaluate the dimensionality and scalability of all MAS subscales including the UL-MAS test items. These investigators, however, provided little demographic information regarding their stroke participants; therefore the extent to which their findings can be generalized to the application of the UL-MAS in individuals with acute/subacute stroke is uncertain. In addition, they did not examine the validity of the difficulty hierarchy of test items UL-MAS, and they were unable to examine item bias or DIF due to the limited demographic information regarding their sample.

Sabari et al. (12) applied Rasch analysis with the primary aim of evaluating the validity of the difficulty hierarchy of the test items within each of the UL-MAS subscales in 100 participants who were a mean of 104 days (range 3 days–6.5 years) post-stroke. These investigators reported inconsistencies in the test items hierarchy and substantial ceiling (28%) and floor (31%) effects in 2 of the 3 subscales. However, the extent to which their calibration of the subscale item difficulties and participant abilities can be confidently applied to the clinical population in which the UL-MAS is generally administered (the acute/subacute phase post-stroke) is uncertain. Their participants had had their strokes as much as 6 years prior to participating in the study and 44–54% of participants were excluded because they had extreme scores (scored 0 or 6 for all test items) leaving data from fewer than 55 cases. Finally, no analysis of item response bias (DIF) was reported.

The current investigation was undertaken as part of a larger study of participants in the acute/subacute phase post-stroke. The aim was to use Rasch analysis to assess the fit of the UL-MAS items, test for potential item bias, check the dimensionality and targeting of scale, and to determine the item difficulty hierarchy.

METHODS

Participants

Ethics approval was obtained from the institutional Human Research Ethics Committees. To be eligible for inclusion in the study, the participants must have experienced their most recent stroke event less than 200 days prior to participating in the study, resulting in unilateral UL impairment. Data were obtained from a total of 80 participants who had a stroke a mean of 64.8 days (standard deviation (SD) 53.3) (range 4–193 days) prior to undertaking the study (Table I). The participants were recruited from 2 different sources. Thirty participants were assessed on 3 occasions as part of a separate clinical trial; while an additional 50 in- and out-patients with stroke were assessed on a single occasion specifically for the present study. Twelve of the 80 participants (15%) had a history of previous stroke(s). A small proportion of the participants had had a haemorrhage. Forty-eight participants had had cortical strokes, while

Table I. *Participant characteristics (n = 80)*

Demographics	
Age (years): mean (SD)	67.4 (15.6)
range	28–90
Gender, *n* (%)	
Male	46 (57.5)
Female	34 (42.5)
Dominant upper limb*, *n* (%)	
Right	78 (97.5)
Left	2 (2.5)
Stroke type, *n* (%)	
Ischaemic	64 (80.0)
Haemorrhagic	15 (18.8)
Other	1 (1.2)
Stroke location, *n* (%)	
Cortical	48 (60.0)
Subcortical	29 (36.2)
Undetermined	3 (3.8)
Stroke side, *n* (%)	
Right	34 (42.5)
Left	46 (37.5)
Affected upper limb, *n* (%)	
Dominant	36 (45.0)
Non-dominant	44 (55.0)

*Hand dominance was determined using the Edinburgh Handedness Questionnaire (13).
SD: standard deviation.

the lesions of 29 participants were located in the subcortical regions of the brain. Three participants had not undergone imaging and therefore the location of their lesions could not be confirmed.

Data collection

The 30 participants in the separate clinical trial were appraised by a trained assessor on 3 occasions, providing a total of 90 observations. These participants were assessed within 6 weeks of stroke onset (mean 27.3 days (SD 9.4) ; range 8–41 days), on a second occasion a mean of 57.8 days post-stroke (SD 10.3) (range 35–74 days) and, finally, at a third time-point a mean of 159.2 days post-stroke (SD 16.0) (range 129–193 days). Additional data were collected from 50 participants with stroke recruited from 3 rehabilitation facilities in the metropolitan Melbourne area. Participating physiotherapy departments were provided with an orientation and a review session to familiarize staff with the data collection form and standardized procedures for administering the UL-MAS. These participants were assessed a mean of 34.4 (SD 38.5) days post-stroke (range 4–178 days). In total 140 sets of scores were obtained. Although there are no clear guidelines concerning the sample size required for Rasch analysis, Linacre (14) suggests that if a scale is well targeted then a sample size of 108 will give 99% confidence of the person estimate being within ± 0.5 logits. The sample size of 140 observations exceeded this value.

Participants were assessed using the 3 UL-MAS subscales: MAS 6: Upper Arm Function, MAS 7: Hand Movements and MAS 8: Advanced Hand Activities. Each subscale consists of 6 test items assessing a range of UL activities along the International Classification of Functioning, Disability and Health (ICF) impairment-activity limitation continuum (15) (Table II). All test items in the MAS 6, and the first 3 and final test items in the MAS 7 subscale reflect the impairments of the "Body Functions" domain of the ICF, specifically "control of voluntary movements (b760)" (16) The remaining 2 test items in the MAS 7 and test items in the MAS 8 subscale evaluate activities in the "Carrying, moving and handling objects (d430–449)" category of "Mobility (d4)", and simulated activities reflecting "Eating (d550)", Drinking (d 560)" and "Caring for body parts (d520)" categories of "Self-care (d5)" within the ICF "Activities and Participation" domain (16).

Table II. *Data collection sheet for Rasch Modelling of the upper limb subscales of the Motor Assessment Scale*

MAS 6: UPPER ARM	1	0
61. Supine lying, protract shoulder girdle with upper arm in elevation		
62. Supine lying, hold extended upper arm in elevation for 2 s		
63. Supine lying, flexion & extension of elbow to take palm to forehead with arm as in 62		
64. Sitting, hold extended arm in forward flexion at 90° to body for 2 s		
65. Sitting, patient lifts arm to above position & holds it there for 10 s, then lowers it		
66. Standing, hand against the wall. Maintain arm position while turning body towards wall		

MAS 7: HAND MOVEMENTS	1	0
71. Sitting, extension of wrist (have patient sitting at table with forearms resting on the table. Place cylindrical object in palm of patient's hand)		
72. Sitting, radial deviation of wrist. (Place forearm in mid-position, i.e. rest on ulnar side, thumb in line with forearm & wrist in extension. place cylindrical object in patient's hand)		
73. Sitting, elbow into side, pronation & supination (3/4 ROM is acceptable)		
74. Reach forward, pick up a large ball (14 cm) with both hands & put it down (must place ball far enough in front to require full extension)		
75. Pick up a polystyrene cup from the table & put it down on the table across the other side of the body (no alteration in shape of cup allowed)		
76. Continuous alternating opposition of thumb & each finger more than 14× in 10 s		

MAS 8: ADVANCED HAND ACTIVITIES	1	0
81. Picking up the top of a pen & putting it down again (arms length to close to body)		
82. Picking up one jelly bean from a cup & placing it in another cup (arms length, L takes bean from R & releases in L cup)		
83. Drawing horizontal lines to stop at a vertical line 10× in 20 s (5 lines must be accurate)		
84. Holding a pencil, make rapid consecutive dots on a sheet of paper (must position pencil in hand w/o assist. approximately, 2 dots/s for 5 s, strokes not allowed)		
85. Take a dessert spoon of liquid to mouth (no head lowering & no spillage)		
86. Holding a comb & combing hair at back of head (no head lowering allowed)		

1: participant able to execute item according to stated performance criteria; 0: participant unable to execute item according to stated performance criteria.

The test items within each subscale are assumed to be hierarchically organized by difficulty. Once an individual is unable to perform a test item it is assumed that they cannot perform any of the remaining test items, thereby reducing the burden of assessment for the therapist and the patient. A single ordinal score between 0 (unable to perform test item 1) to 6 (able to meet performance criteria for all 6 test items) is recorded for each subscale. For the purposes of this study, administration of the UL-MAS was undertaken according to published scoring criteria and guidelines (3). However, to investigate the test item difficulty hierarchy, all test items were administered to the participant. Each test item was scored using a dichotomous response ("0" = unable to execute item, "1" = able to execute item according to stated performance criteria) (Table II). Factor analysis had previously shown the composite score of the 3 UL-MAS subscales items could be validly applied to quantify motor recovery of the whole UL following stroke (17); hence all analyses were undertaken using the composite UL-MAS.

Data analysis

The RUMM2020 Rasch measurement software program (Version 4.0 1997–2004, RUMM Laboratory Pty Ltd) was used to perform the Rasch analyses of the combined subscales of the UL-MAS. Following the procedures recommended by Wright (18) the data set was "stacked" including multiple measurements from participants who were assessed across a number of time-points. This procedure can introduce some dependency into the data, but this is likely to be minor (18) and "stacking" provides the advantage of allowing a formal assessment of the invariance of test items over time; an important issue when tools are used to assess change.

All data analysis procedures undertaken followed the recommendations of Pallant & Tennant (19) and Hagquist et al. (20). The overall fit of the UL-MAS scale was evaluated using χ^2 statistics with non-significant χ^2 probability values (after Bonferroni adjustment for number of items) indicating good fit to the Rasch model (21). Standardized fit residual statistics were also examined, with good fit indicated by a mean score of 0.0 and a SD of 1.0 (22). The fit of individual persons and items was assessed with non-significant χ^2 probability values, and standardized fit residuals of between –2.5 and +2.5 indicating adequate fit (22).

Response dependency between test items was appraised by inspecting the residual correlation matrix for pairs of items with correlation values greater than 0.3 (23). The violation of the assumption of local independence among the items can artificially inflate reliability measures (24, 25). Where high residual correlations were detected these items were combined into "testlets" and the reliability estimate compared with the original value obtained. A substantial drop in the PSI or Cronbach's alpha value would indicate local dependency among the items (20) and a potential loss of discriminative capacity for individual application (10).

Uni-dimensionality of the UL-MAS was assessed by performing principal component analyses to identify subsets of items with positive and negative loadings on the first unrotated component. Rasch-derived person estimates from these subsets were compared using a series of *t*-tests. If more than 5% of these tests are significant (or specifically the lower bound of the binomial confidence interval is above 5%), the scale is deemed to be multidimensional (25).

The presence of item response bias (DIF) was evaluated using analysis of variance for gender, age (≤64 years or over 64 years), stroke type (infarct or haemorrhage), general location of the current stroke lesion (cortical or subcortical), and history of previous stroke incident (s) (yes or no). Differential item functioning was also assessed for time since the admission stroke incident (26). This was done to assess the invariance of scores repeated over multiple time-points (27). The time-points were chosen to represent the approximate time-points at which participants in the separate clinical trial were assessed (≤42 days, 43–90 days, and 91–200 days).

After a satisfactory solution was obtained the order of difficulty of the items was checked against the original UL-MAS test order to establish the validity of the difficulty hierarchy of the scale.

The targeting of the UL-MAS as applied to the acute/subacute stroke participants was examined by: (*i*) calculating mean participant ability in relation to overall difficulty of the UL-MAS test items; (*ii*) calculating the standard error between test items to establish the "spread" of the item difficulties as a measure of the precision of the scale; (*iii*) calculating the percentage of participants that were unable to execute any of the test items (floor effects) or capable of completing all test items (ceiling effects). In addition, the Person Separation Index (PSI) calculated by the RUMM2020 programme was applied to estimate the

capacity of the UL-MAS to distinguish between or stratify groups of stroke participants with differing UL abilities (10).

RESULTS

Initial fit of the data to the Rasch model

The UL-MAS showed good fit to the Rasch model ($\chi^2_{(1,17)} = 22.648$; $p = 0.204$), with no evidence of misfitting items (mean –0.473, SD 0.566) or persons (mean –0.272, SD 0.371). No significant item response bias (DIF) was detected based on the gender, stroke type, stroke location, stroke history of the participants. Responses were found to be invariant across time and repeated assessments; supporting the decision to combine the data collected at the 3 different time-points. Significant DIF was, however, detected for age for test item 72 (radial deviation of the wrist; $F_{(1,109)} = 9.297$; $p = 0.003$). This test item was found to be systematically easier for participants under 65 years of age to perform (–3.291 logits) compared with participants 65 years of age and older with similar abilities (–0.734 logits). It was therefore decided to delete item 72 from the UL-MAS and to test the fit of the remaining 17 items (Table III).

Rasch analysis of the 17-item UL-MAS

Rasch analysis of the 17 remaining test items in the UL-MAS revealed good fit to the model ($\chi^2_{(1,16)} = 20.451$; $p = 0.252$), excellent reliability (PSI = 0.96), no evidence of misfitting items (mean –0.412, SD 0.461) or persons (–0.234, SD 0.314), and no DIF for any of the tested factors.

Inspection of the residual correlation matrix revealed a number of pairs of items with correlation coefficients above 0.3 (refer to Table II for item descriptions): test items 64 and 65 (0.58), items 75 and 76 (0.38), and items 81 and 82 (0.50). To test the impact of these breaches of local independence on the reliability estimates these pairs of items were combined

into "testlets". There was only a very small drop in reliability estimates after combining these items (PSI = 0.95, Cronbach's alpha = 0.93), as compared with the original estimates (PSI = 0.96, Cronbach's alpha = 0.95), indicating no serious breach of the assumption of local independence of the items.

Principal component analysis confirmed the uni-dimensionality of the scale. Differences in scores between positive and negative loading test items resulted in significant t-tests ($p < 0.05$) for only 2.85% of participants, which fell well below the acceptable guideline of 5%.

Test item ordering according to estimated difficulty (or item locations) on the logits scale is presented in Table IV. The validity of the difficulty hierarchies for the test items in the MAS 6 Upper Arm and MAS 7 Hand Movements subscales were confirmed. Inconsistencies in the ordering of the test items by difficulty were identified in the MAS 8 Advanced Hand Activities subscale (Table IV). Test items 83 (drawing horizontal lines) and 84 (making consecutive dots with a pencil) were found to be more difficult for the participants to successfully execute than items 85 (bringing a spoonful of liquid to mouth) and 86 (combing hair at back of head; Table II).

The targeting of the UL-MAS test items in relation to the abilities of the 140 participant observations on the same generic logits scale are presented in Fig. 1. The mean person ability estimate for the participants with acute/subacute stroke was –1.21 logits (SD 3.97), suggesting that their abilities were generally less than the difficulty of the test items. Nine percent of participants successfully performed all test items (ceiling effects), while 14% were unable to execute any of the test items (floor effects). The spread of difficulty of the test items is illustrated by the blocks below the horizontal axis in Fig. 1 and item difficulty and standard error values in Table IV. There is evidence of relatively large gaps in the estimated difficulties of some sequential test items (e.g. a difference of over 5 standard errors separates items 61 and 62), while

Table III. *Initial tests of fit and individual test item difficulty for the Upper Limb subscales of the Motor Assessment Scale*

Test item	Difficulty (logits)	Standard error (logits)	Standardized fit residuals	χ^2 statistic	χ^2 probability	Frequency of participants scoring 0	Frequency of participants scoring 1
61	–5.512	0.396	–0.221	1.324	0.249	6	107
62	–3.406	0.327	–0.770	1.361	0.243	29	84
63	–2.750	0.327	–0.867	1.488	0.222	34	79
64	–0.962	0.309	0.119	0.499	0.480	48	65
65	0.280	0.295	0.011	2.423	0.119	61	52
66	3.719	0.364	0.200	3.921	0.048	102	11
71	–2.821	0.327	–0.402	0.079	0.779	33	80
72	–1.630	0.319	0.432	1.214	0.270	43	70
73	–2.067	0.324	–0.254	1.168	0.279	41	72
74	–0.577	0.304	–0.984	0.629	0.427	54	59
75	0.973	0.292	–0.735	1.376	0.240	73	40
76	2.599	0.312	–0.196	0.365	0.546	91	22
81	–0.211	0.300	–1.813	2.312	0.128	60	53
82	0.521	0.294	–1.430	1.683	0.194	67	46
83	4.102	0.393	–0.176	0.643	0.422	105	8-
84	2.858	0.321	–0.555	0.506	0.477	93	20
85	2.366	0.306	–0.374	0.841	0.359	88	25
86	2.513	0.309	–0.497	0.815	0.366	89	24

All χ^2 statistics were undertaken with 1 degree of freedom (Bonferroni correction, $p = 0.003$).

Table IV. *The final fit and ordering of individual item difficulty for the 17-item revised Upper Limb subscales of the Motor Assessment Scale*

Test item	Difficulty (logits)	Standard error (logits)	Standardized fit residuals	χ^2 statistic	χ^2 probability	Frequency of participants scoring 0	Frequency of participants scoring 1
61	−6.165	0.426	−0.201	1.010	0.315	5	107
62	−3.791	0.347	−0.722	1.433	0.231	28	84
63	−2.982	0.350	−0.692	1.249	0.263	33	79
71	−2.979	0.350	0.018	0.047	0.827	32	80
73	−2.229	0.345	−0.086	0.316	0.573	40	72
64	−0.965	0.320	0.078	0.779	0.377	47	65
74	−0.598	0.313	−0.817	0.258	0.611	53	59
81	−0.241	0.306	−1.473	2.280	0.131	59	53
65	0.338	0.298	−0.066	1.723	0.189	60	52
82	0.536	0.296	−1.216	2.400	0.121	66	46
75	0.971	0.293	−0.622	0.807	0.369	72	40
85	2.361	0.304	−0.298	0.659	0.419	87	25
86	2.487	0.307	−0.393	0.610	0.435	88	24
76	2.599	0.311	−0.209	0.689	0.407	90	22
84	2.835	0.318	−0.444	0.403	0.526	92	20
66	3.735	0.365	0.118	5.112	0.023	101	11
83	4.088	0.391	−0.142	0.676	0.411	104	8

All χ^2 statistics were undertaken with 1 degree of freedom (Bonferroni correction, $p = 0.003$).

relatively similar levels of difficulty are observed between other sequential test items (e.g. items 63 and 71, items 85 and 86).

DISCUSSION

The present study confirms the results of previously published research in finding the UL-MAS to be a uni-dimensional scale

that measures a single construct, UL motor recovery (11, 12, 16). After removal of item 72, which showed DIF for age, the remaining 17 test items showed adequate fit to the Rasch model, with good internal consistency. Contrary to the findings of Sabari et al. (12), the present study validates the test item difficulty hierarchy in the MAS 6 Upper Arm and MAS 7 Hand Movements subtests. As previously reported, test items in MAS
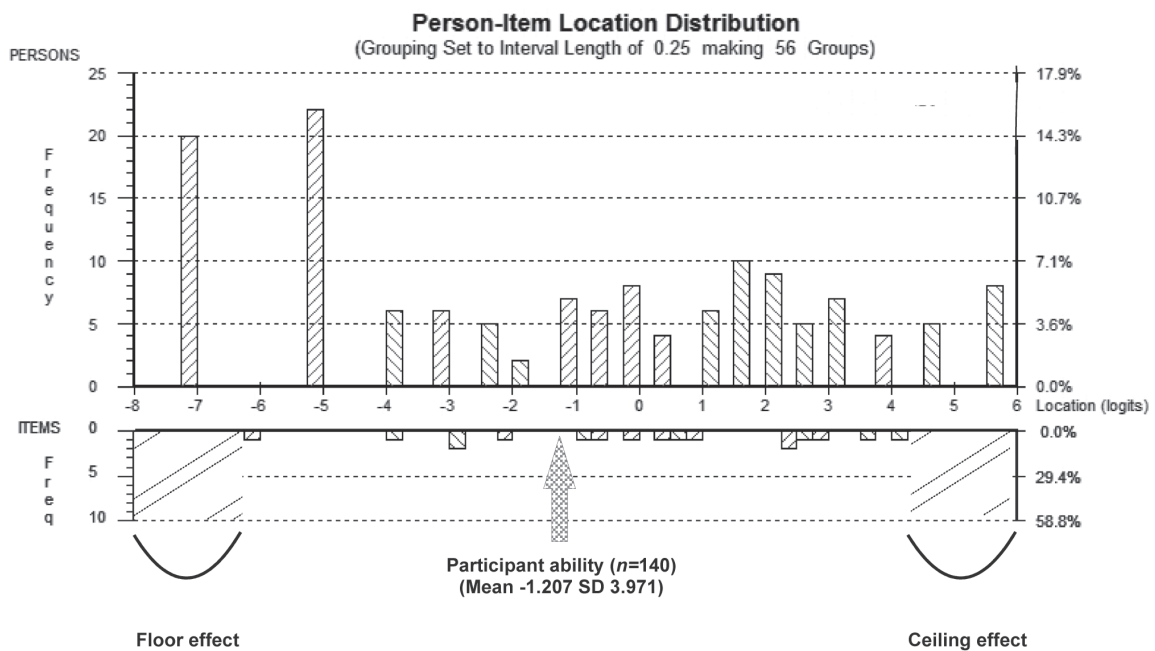

Fig. 1. Targeting of the Upper Limb subscales of the Motor Assessment Scale (UL-MAS) test items to the abilities of the participants with acute/subacute stroke. The upper columns represent the number of participants at each ability level on the logits scale, while the columns at the bottom of the figure represent the number of test items at each item difficulty level on the same scale. Negative logit values indicate test items are easier to perform/participants have less UL motor recovery. Positive logit values indicate test items are more difficult to perform/participants have greater UL motor recovery. The mean participant ability is indicated with a large shaded arrow on the scale. Clustering of participants unable to perform any UL-MAS test items (floor effects) or able to perform all test items (ceiling effects) are represented on each side of the scale.

8 were not ordered in accordance with their estimated difficulty (12, 28, 29). Evidence from the present study supports the psychometric properties of the revised 17-item UL-MAS in individuals in the acute/subacute phase post-stroke. The scale demonstrates acceptable ceiling and floor effects, and an excellent capacity to stratify participant groups in this clinical population on the basis of their UL motor recovery.

Rasch analysis has previously been applied to evaluate psychometric properties of the UL-MAS; however, the present study is the first to investigate DIF related to the demographic and stroke-related characteristics of the participants. Item response biases were not found related to the gender, stroke type, stroke location, time since stroke or time-frame of assessment. Significant DIF was, however, detected for test item 72 (radial deviation of the wrist) based on the participant age group (≤64 years or 65+ years). It is uncertain why radial deviation of the wrist was uniformly more difficult to perform for the older participants. It could be hypothesized that active range of motion at the wrist might be reduced in older individuals, and they might therefore appear to have reduced active radial deviation of the wrist. However, there was a trend for observed performance on the test item to be inconsistent with expected performance even within the younger age group of stroke participants. These findings suggest factors other than upper limb motor recovery bias performance of this test item. Deletion of item 72 improved the "fit" of the UL-MAS to the Rasch model.

Evidence suggests that decisions regarding deletion of misfitting test items should be taken together with knowledge of the construct validity of the specific item and the measurement scale (30). When the UL-MAS were developed by Carr et al. (3), radial deviation of the wrist combined with wrist extension were identified as "essential movement components" required for function in grasp and release. A more recent publication by these investigators no longer advocates radial deviation of the wrist as an essential movement to be retrained for reaching and grasping (31). There is kinematic evidence of radial deviation when the trajectory of the hand is adjusted in reaching toward an object (32); however, this movement is thought to be co-ordinated as part of the total synergy (33) or programme of movement (34), rather than occurring in isolation as tested in the UL-MAS. Therefore, the construct validity of test item 72 is no longer substantiated by current literature. This evidence, taken together with findings of the current Rasch analysis, suggests that item 72 does not add meaningfully to the assessment of motor recovery in the UL-MAS.

DIF analysis was also used in this study to assess the invariance of item response over multiple time-points, an important issue for clinicians and researchers wishing to use the UL-MAS to measure change over time. Observations from 3 time periods were included, with the data "stacked" according to the guidelines recommended by Wright (18). No DIF was detected for time-point, supporting the invariance of the scale over time, and justifying the comparison of UL-MAS scores across time-points to assess change in UL motor recovery. However, the effect of the potential person dependency upon estimates associated with the repeated measures in some of the study

participants is not known and requires further methodological investigation.

While local item dependency could have been problematic due to the original design and scoring of the instrument, only 3 sets of item pairs showed elevated item correlations. Assessment of these item pairs as testlets indicated that these items did not result in any artificial inflation of the reliability estimates, supporting their inclusion in the UL-MAS (20). The level of correlation between the items was not so high as to suggest item redundancy or indicate the removal of items was warranted.

In agreement with previous studies, the UL-MAS test items that required skilful use of a pen or pencil on paper were found to be the most difficult tasks for participants with stroke. The test items within the MAS 8 Advanced Hand Activities subscale were not found to be ordered with respect to hierarchy of difficulty (12, 28, 29). As recommended previously, clinicians should administer all test items within this subscale to establish the highest level of hand function their patients with stroke are capable of achieving (28, 29).

Contrary to the findings reported by Sabari et al. (12), test items in subtests MAS 6 Upper Arm and MAS 7 Hand Movements were found to be ordered with respect to their difficulty. The difference in findings between the present study and the findings reported by Sabari et al. (12) is potentially attributable to the number of observations used to estimate the test item difficulties. The present study examined the composite of the 3 MAS subscales of arm and hand function in participants who were relatively early in their recovery post-stroke (mean 67 days from stroke onset). Twenty percent of 140 observations were excluded from analysis as extreme scores. Item difficulty estimates in the present study were based on 112 observations, and as previously discussed, in accordance to recommendations made by Linacre (35) there were a minimum of 10 observations in all scale categories with the exception of the most difficult test item (Table IV). Sabari et al. (12) analysed each subscale separately in participants much further along in their post-stroke recovery (mean of 104 days from stroke onset). As a result, a larger proportion of participant observations were classified as extreme scores (44–54%), and estimates of item difficulty were based on half as many observations (55 observations) and smaller numbers of observations within each scale category, leading to less robust estimates for difficulties of the test items.

The present study indicates that the targeting of the revised 17-item UL-MAS was appropriate for participants with acute/subacute stroke. Floor and ceiling effects can differ depending upon the characteristics of the participants and whether UL-MAS scores are assigned for individual subtests or as a composite score. Floor effects as high as 58% have been reported when the UL-MAS was administered to 48 inpatients with a median of 24 days post-stroke (36), and ceiling effects of 39% have been reported in a retrospective chart audit of 153 patients discharged at a non-specific length of time post-stroke from rehabilitation (37). In the present study, the floor (14%) and ceiling (9%) effects of the composite UL-MAS score in participants 14–200 days post-stroke were found to

be well within a range indicative of a suitable measurement model (38).

Evaluation of the spread of difficulty of test items within a scale highlights the balance between the comprehensiveness and precision of the measurement scale, and the clinical utility and burden of assessment in patients after stroke (39). When the spread of the difficulty of test items of the UL-MAS was examined, relatively large gaps of greater than 2 standard errors were found between some sequential test items. This finding was in agreement with the results of Sabari et al. (12). Nonetheless, the high Person Separation Index (0.96) for the composite scale implies that the scale has a suitable pool of test items so that differences in ability can be differentiated or stratified (10). This reliability value suggests that the 17 item-UL-MAS scale has the precision to stratify participants with stroke on the basis of UL motor recovery into more than 4 different strata (40), providing support for the use of the UL-MAS in evaluating the effectiveness of UL interventions in a clinical trial.

It must be acknowledged, however, that these "gaps" potentially diminish the sensitivity of the UL-MAS to individual changes in UL ability. This limitation has implications for clinical settings where therapists would use the scale in patients with stroke to evaluate change over time. English et al. (41) have reported that the individual UL subscales showed relatively small effect sizes in inpatients with stroke. It is unknown if a larger effect size might have been obtained using a composite UL-MAS score that reflected total UL activity limitation. Sabari et al. (12) have made recommendations regarding the inclusion of additional test items to "fill" the apparent gaps in difficulty between sequential items in the subscales. These recommendations remain to be explored.

In the current study test items were also found to have item locations that were very close together on the logit scale. This finding could be interpreted as an indication of redundancy of items within the scale, however construct validity and the purposes of the scale must also be considered before items are deleted (30). For example, the difficulties of item 85 (bringing spoonful of liquid to mouth) and item 86 (combing hair at back of head) were within 1 standard error of each other. While it could be argued that removing one of these items could provide similar information about the ability level of the participants while reducing burden of administration, each item evaluates different combinations of UL movements and relevant everyday tasks potentially useful to goal setting and treatment planning. Therefore, there is a requirement to balance the need for clinically meaningful data from the requirements of measurement criteria. In addition, given the findings of the present study are based on a relatively modest number of observations, it would be prudent for a future study to re-examine these findings using a larger clinical sample.

In summary, the 3 subscales comprising the UL-MAS have been shown to be a uni-dimensional scale offering a measurement tool that is appropriately targeted for the assessment of UL motor recovery in acute/subacute stroke participants within 200 days of their stroke. Evidence from the present study suggests that the 17-item revised version of the scale has the

precision to distinguish or stratify groups of participants with differing UL abilities following stroke, supporting the choice of the composite UL-MAS as a potential outcome measure for use in clinical trials to evaluate the effectiveness of interventions used to improve UL motor abilities post-stroke. Based on these findings it is recommended that researchers and clinicians assess all items in 3 subscales of the UL-MAS in their participants or patients after stroke.

ACKNOWLEDGEMENTS

REFERENCES

1. Nakayama H, Jorgensen HS, Raaschou HO, Olsen TS. Recovery of upper extremity function in stroke patients: the Copenhagen Stroke Study. Arch Phys Med Rehabil 1994; 75: 394–398.
2. Lai SM, Studenski S, Duncan PW, Perera S. Persisting consequences of stroke measured by the Stroke Impact Scale. Stroke 2002; 33: 1840–1844.
3. Carr JH, Shepherd RB, Nordholm L, Lynn D. Investigation of a new motor assessment scale for stroke patients. Phys Ther 1985; 65: 175–179.
4. Loewen SC, Anderson BA. Reliability of the Modified Motor Assessment Scale and the Barthel Index. Phys Ther 1988; 68: 1077–1081.
5. Kjendahl A, Jahnsen R, Aamodt G. Motor assessment scale in Norway: translation and inter-rater reliability. Adv Physiother 2005; 7: 7–12.
6. Wright BD, Linacre JM. Observations are always ordinal: measurement, however, must be interval. Arch Phys Med Rehabil 1989; 70: 857–860.
7. Wright BD, Masters GN. Rating scale analysis. Chicago: MESA Press; 1982.
8. de Koning CHP, van den Heuvel SP, Staal JB, Smits-Engelsman BCM, Hendriks EJM. Clinimetric evaluation of active range of motion measures in patients with non-specific neck pain: a systematic review. Eur Spine J 2008; 17: 905–921.
9. Andrich D. Rasch Models for Measurement. California: Sage Publications Inc.; 1988.
10. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. New Jersey: Lawrence Erlbaum Associates Publishers; 2001.
11. Aamodt G, Kjendahl A, Jahnsen R. Dimensionality and scalability of the Motor Assessment Scale (MAS). Disabil Rehabil 2006; 28: 1007–1013.
12. Sabari JS, Lim AL, Velozo CA, Lehman L, Kieran O, Lai JS. Assessing arm and hand function after stroke: a validity test of the hierarchical scoring system used in the motor assessment scale for stroke. Arch Phys Med Rehabil 2005; 86: 1609–1615.
13. Bryden MP. Measuring handedness with questionnaires. Neuropsychologia 1977; 15: 617–624.
14. Linacre JM. Sample size and item calibration stability. Rasch Measurement Transactions. [serial on the Internet, cited 2009 Sept 1] 1994; 7: 328. Available from: http://www.rasch.org/rmt/rmt74m.htm
15. World Health Organization. Towards common language for functioning, disability and health (ICF). Geneva; 2002 [cited 2006 May 6]. Available from: http://www3.who.int/icf/icftemplate.cfm?myurl=beginners.html&mytitle=Beginner%27s%20Guide

16. World Health Organization. International Classification of Functioning, Disability and Health (ICF) (On-line version). Geneva 2007 [cited 2009 Sept 18]. Available from: http://apps.who.int/classifications/icfbrowser/

17. Lannin N. Reliability, validity and factor structure of the upper limb subscale of the Motor Assessment Scale (UL-MAS) in adults following stroke. Disabil Rehabil 2004; 26: 109–116.

18. Wright BD. Rack and stack: time 1 vs. time 2. Rasch Measurement Transactions [serial on the Internet, cited 2009 Sept 9] 2003; 17: 905–906. Available from: http://www.rasch.org/rmt/rmt171.htm

19. Pallant J, 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.

20. Hagquist CA, Bruce MJ, Gustavsson P. Using the Rasch model in nursing research: an introduction and illustrative example. Int J Nurs Stud 2009; 46: 380–393.

21. Bhakta B, Lawton G, Andrich D, Tennant A, Horton M. Using item response theory to explore the psychometric properties of extended matching questions examination in undergraduate medical education. BMC Med Educ [serial on the Internet, cited 2007 June 22] 2005; 5: 9. Available from: http://www.biomedcentral.com/1472–6920/5/9

22. Linacre JM. What do infit and outfit, mean-square and standardized mean? Rasch Measurement Transactions [serial on the Internet, cited 2007 June 22] 2002; 16: 878. Available from: http://www.rasch.org/rmt/rmt162f.htm

23. Smith EVJ. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3: 205–231.

24. Marais I, Andrich D. Formalizing dimension and response violations of local independence in the unidimensional Rasch model. J Appl Meas 2008; 9: 200–215.

25. Tennant A, Conaghan P. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007; 57: 1358–1362.

26. Chang WC, Chan C. Rasch analysis for outcomes measures: some methodological considerations. Arch Phys Med Rehabil 1995; 76: 934–939.

27. Woodbury ML, Velozo CA, Richards C, Duncan PW, Studenski S, Lai SM. Longitudinal stability of the Fugl-Meyer Assessment of the upper extremity. Arch Phys Med Rehabil 2008; 89: 1563–1569.

28. Dean CM, Mackey F. Motor assessment scale scores as a measure of rehabilitation outcome following stroke. Aust J Physiother 1992; 38: 31–35.

29. Poole JL, Whitney SL. Motor assessment scale for stroke patients: concurrent validity and interrater reliability. Arch Phys Med Rehabil 1988; 69: 195–197.

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

31. Carr JH, Shepherd RB. Stroke Rehabilitation: guidelines for exercise and training to optimize motor skill. London: Butterworth-Heinemann; 2003.

32. Fan J, He J, Helms Tillery SI. Control of hand orientation and arm movement during reach and grasp. Exp Brain Res 2006; 171: 283–296.

33. Mason CR, Gomez JE, Ebner TJ. Hand synergies during reach-to-grasp. J Neurophys 2001; 86: 2896–2910.

34. Marotta JJ, Medendorp WP, Crawford JD. Kinematic rules for upper and lower arm contributions to grasp orientation. J Neurophys 2003; 90: 3816–3827.

35. Linacre JM. Investigating rating scale category utility. J Outcome Meas 1999 1999; 3: 103–122.

36. Hsueh IP, Hsieh CL. Responsiveness of two upper extremity function instruments for stroke inpatients receiving rehabilitation. Clin Rehabil 2002; 16: 617–624.

37. Williams BK, Galea MP, Winter AT. What is the functional outcome for the upper limb after stroke? Aust J Physiother 2001; 47: 19–27.

38. Andresen EM, Rothenberg BM, Panzer R, Katz P, McDermott MP. Selecting a generic measure of health-related quality of life for use among older adults: a comparison of candidate instruments. Eval Health Prof 1998; 21: 244–264.

39. Jette AM, Haley SM. Contemporary measurement techniques for rehabilitation outcome assessment. J Rehabil Med 2005; 37: 339–345.

40. Fisher WJ. Reliability statistics. Rasch Measurement Transactions. [serial on the Internet, cited 2007 June 22] 1992; 6: 238. Available from: http://www.rasch.org/rmt/rmt63i.htm

41. English CK, Hillier SL, Stiller K, Warden-Flood A. The sensitivity of three commonly used outcome measures to detect change amongst patients receiving inpatient rehabilitation following stroke. Clin Rehabil 2006; 20: 52–55.