J Rehabil Med 2009; 41: 1055–1061

ORIGINAL REPORT

PSYCHOMETRIC PROPERTIES OF THE RIVERMEAD MOTOR ASSESSMENT: ITS UTILITY IN STROKE

Yeşim Kurtaiş, MD[1], Ayşe Küçükdeveci, MD[1], Atilla Elhan, PhD[2], Ahmet Yılmaz, MD[1], Tuğba Kalli, MD[1], Birkan Sonel Tur, MD[1] and Alan Tennant, PhD[3]

From the [1]Department of Physical Medicine & Rehabilitation and [2]Department of Biostatistics, Ankara University Faculty of Medicine, Ankara, Turkey and [3]The University of Leeds, Faculty of Medicine and Health, The General Infirmary at Leeds, Leeds, UK

Objective: **To investigate the psychometric properties of the Rivermead Motor Assessment by Rasch analysis and conventional statistics to improve its clinical utility.**
Methods: **A total of 107 patients after stroke were evaluated using the Rivermead Motor Assessment and Functional Independence Measure (FIM™). Scaling properties were assessed using Mokken scaling, internal construct validity using Rasch analysis, reliability using Cronbach's alpha and intra-class correlation coefficients, external construct validity through convergent validity with FIM™, and responsiveness using the effect size and standardized response mean.**
Results: **Cronbach's alpha and intra-class correlation coefficients for 3 sections of the Rivermead Motor Assessment were between 0.88 and 0.95. Mokken scaling showed appropriate Guttman patterns, but the hierarchical ordering of the items differed from that of the original. After removing 4 items of *gross function*, 1 of *leg-trunk*, and 4 of *arm,* all sections met Rasch model expectations. External construct validity was confirmed. Mean values of effect size and standardized response were 0.38–0.51 and 0.60–0.89, respectively.**
Conclusion: **The Rivermead Motor Assessment has been shown to be reliable and responsive. Guttman scaling is apparent, but not as originally defined. After removing some items, the scale satisfies the most stringent Rasch measurement criteria and can produce interval scaling for the assessment of motor function in stroke.**

Key words: stroke, motor assessment, reproducibility of results, rehabilitation, outcome assessment.

J Rehabil Med 2009; 41: 1055–1061

Correspondence address: Yeşim Kurtaiş, Ibni Sina Hastanesi Fiziksel Tıp ve Rehabilitasyon AD 06100 Sıhhiye Ankara, Turkey. E-mail: kurtais@medicine.ankara.edu.tr, kurtais@gmail.com

Submitted January 26, 2009; accepted August 26, 2009

INTRODUCTION

Professionals working in the field of rehabilitation have traditionally given considerable emphasis to decreasing the limitations in activity of those who have experienced a stroke, despite the existence of various continuing impairments. However, the promising results of an increasing number of studies about neuroplasticity have led rehabilitation specialists to pay increasing attention to neurological recovery including motor function (1, 2). Therefore, among various attributes, the accurate evaluation of motor function becomes important for both the planning of interventions and the assessment of outcome.

The Rivermead Motor Assessment (RMA) is one of the tools that assesses and defines the motor function of stroke patients. It consists of 3 sections: gross function, leg and trunk, and arm. Typically, for the gross function and arm sections, items are organized hierarchically, and where a patient fails 3 consecutive items, the test on that section is stopped. Based on the Bobath therapeutic approach (3), it is commonly used as an outcome measure both in clinical setting and research (4–10). The instrument was reported to have good test-retest and inter-rater reliability (11) and its convergent validity with the Motricity Index has been demonstrated (12). Given the way in which it is commonly applied and scored, the RMA has also been investigated for scalability. The leg and trunk section was found to fail Guttman scaling criteria in acute stroke patients, and only the items in the gross function section met scaling criteria with non-acute strokes (13, 14). This suggests that current assessment practice of stopping after 3 failures may be compromised, as the hierarchical structure of the items do not support such actions, at least in the way originally prescribed by the developers. Furthermore, the lack of conformity to Guttman scaling may suggest problems with the underlying dimensionality of the subscales, so compromising any score used for outcome assessment.

Over the past decade, the modern psychometric approaches of item-response theory have been increasingly applied to the evaluation of the instruments used to measure health outcomes, including the scaling properties of the instrument (15). Given the widespread use of the RMA, and the concerns about its scalability, the construct validity and unidimensionality of the arm section of the scale was recently investigated using Rasch analysis (16). However, the internal construct validity of other 2 sections of RMA was not assessed and consequently this current study sets out to investigate in detail the psychometric properties of all the subscales of the RMA, including their scalability. To attain this goal, Mokken scaling (17) and Rasch analysis were used to investigate the internal construct

doi: 10.2340/16501977-0463
ISSN 1650-1977

validity and scaling properties of all sections of RMA. The external validity of the scale and its sensitivity to change were also assessed by conventional statistical methods.

SUBJECTS AND METHODS

Subjects and settings

The study was conducted in the rehabilitation unit of a university hospital. Consecutive patients admitted to in-patient rehabilitation unit with unilateral hemiplegia and diagnosed as first-ever stroke (18) were recruited to the study between January 2005 and December 2007. Exclusion criteria were: other neurological diseases with permanent damage; other musculoskeletal impairments; apparent cognitive deficits hampering cooperation and understanding or following the instructions for motor testing, and existence of apraxia. The study was approved by the ethics committee of the Ankara University Faculty of Medicine and all the patients gave their informed consent.

Among the 142 patients screened, 107 patients with a mean age of 62.4 (standard deviation (SD) 12.8) years (range 28–85 years, median 65 years) were included in the study and assessed at admission and on discharge. Sixty percent was male and the median time since stroke was 2 months (mean 5.6 (SD 11.2), range 0.5–78 months). Type of stroke was ischaemic in 79% and haemorrhagic in 21%. Forty-eight percent of the group had right-sided hemiplegia.

Evaluation

The demographic and clinical characteristics of the patients were recorded and patients were assessed with the RMA and the Functional Independence Measure (FIM^{TM}) (19, 20). First, the RMA was adapted to the Turkish population using valid guidelines for cross-cultural adaptation (21). Then, 2 assessors (AY, TK) administered the test to patients until they were familiarized with the testing procedure and eliminated any discrepancy between raters. All study subjects were then assessed at admission and on discharge by the same physiatrists (AY, TK) at all times.

The RMA consists of 3 scales including gross function, leg and trunk, and arm sections. Each activity should be carried out independently. As previously stated, all items in 2 of the sections (not the leg and trunk) are supposed to be in a hierarchical order and traditionally stopping rules are applied after 3 failures on any subscale. However, for the purposes of this evaluation, all items were scored on each subscale, so that the hierarchical ordering of each set of items could be fully evaluated. The Gross Function section (RMA-gf) comprises 13 items and assesses mainly mobility and ambulation from sitting to running. The Leg and Trunk section (RMA-lt) comprises 10 items assessing isolated movements of trunk (e.g. rolling to the affected side) and leg (e.g. dorsiflexion of ankle with leg extended while lying). Finally, the Arm section (RMA-a) comprises 15 items assessing isolated movements (e.g. protracting shoulder girdle with arm in elevation while lying) and complex tasks (e.g. “pat-a-cake 7 times in 15 min). All items are scored 1 if the patient performs the activity, otherwise they are scored 0.

Reliability

Reliability of the RMA was initially tested by internal consistency. The internal consistency of an instrument is an estimate of the degree to which its constituent items are interrelated, and is assessed by Cronbach’s alpha coefficient (22). In addition, the intra-class correlation coefficient is calculated (ICC 2,1) (23). Subsequently reliability is further tested by the person separation index (PSI) from the Rasch analysis (see below). Usually a reliability of 0.70 is required for analysis at the group level, and values of 0.85 and higher for individual use (24).

Internal construct validity and scalability

Initially the data from each subscale were subjected to Mokken scaling to determine if there existed a non-parametric probabilistic Guttman-style relationship in the data (17, 25–26) Acceptability of the probabilistic relationship was determined by a Loevinger *H*-coefficient > 0.3 for individual items and the scale as a whole. Mokken scaling works “bottom-up” by starting with the 2 items that have the strongest correlation, and then adding further items which satisfy the Loevinger level given above. An attempt to construct a second (and subsequent) scale is made when there is more than 1 item remaining. Further details of the process can be found elsewhere (27). The process is viewed as a natural starting point for Rasch analysis, as satisfying Mokken scale requirements is a necessary, but not sufficient, condition for satisfying Rasch model expectations (28).

The unidimensional Rasch model asserts that the easier the item the more likely it will be passed, and the more able the person the more likely they will pass an item compared with a less able person (29, 30). It is a parametric procedure, and makes further requirements of the data such that, if data meet model expectations a transformation from ordinal to interval data can be achieved (31). Full details of applying the Rasch model in health outcomes are given elsewhere (32, 33).

Briefly, data are tested against model expectations and to satisfy these, a non-significant deviation from model expectation is expected. Thus χ^2 based fit statistics (overall and for each item) should be non-significant (Bonferroni adjusted). Residual item fit statistics should fall within ± 2.5 (99% confidence). Overall item and person residuals should display a mean of zero and SD of 1 for perfect fit. A PSI is used to calculate reliability, interpreted in the same way as Cronbach’s alpha (22). There should also be an absence of differential item functioning (DIF), which shows whether the response varies by group membership (e.g. age) at a given level of the construct (34). For DIF, *conditioned upon the patient’s level of motor function*, no difference should be observed in the level of response to a given item by external contextual factors such as age, gender, hemiplegic side, type of stroke and time (admission/discharge).

The assumption of local dependency of items is examined through the correlations of the residuals. Where locally dependent items are observed they are grouped into a testlet (items are added together) to determine if fit is improved. Finally, a formal test of the assumption of unidimensionality is undertaken by performing a principal component analysis of the residuals. Items with the highest positive and negative correlations on the first residual factor are used to construct 2 smaller scales, anchored to the item difficulties of the main analysis (35). The person estimates derived from these 2 subsets of items are contrasted for each individual by an independent *t*-test. A significant difference would be expected to occur by chance in 5% of cases. Consequently the percentage of tests outside the range ± 1.96 is reported, together with a 95% binomial confidence interval. This interval should overlap 5% for a non-significant finding to confirm unidimensionality.

External construct validity

The external construct validity of the RMA is assessed through convergent validity with the FIM^{TM} motor scale, previously adapted for use in Turkey (20). Although a measure of activity limitation rather than motor impairment, a moderate association would be expected. The FIM^{TM} is used as part of a routine clinical procedure for the assessment of patients with neurological disabilities in the present setting. Degree of associations with FIM^{TM} motor and its sub-components, FIM^{TM} self-care and FIM^{TM} mobility (transfer and locomotion) are analysed by Spearman’s correlation coefficient.

Responsiveness

Responsiveness is defined as the ability of an instrument to detect important changes over time. In this study, it is evaluated through the effect size (ES) and standardized response mean (SRM). SRM is calculated by dividing the mean change by the SD of the change scores, while ES is calculated by dividing the mean change by the SD of the admission scores (36).

Sample size and statistical analysis

For the Rasch analysis, it is reported that a sample size of 150 patients will estimate item difficulty, with α of 0.01, to within ± 0.5 logits (37).

Table I. *Frequency (%) of positive responses (hierarchical ordering) within each Rivermead Motor Assessment subscale*

Item number	Gross motor function	Arm	Leg and trunk
1	13.4	16.4	14.2
2	12.1	10.9	12.9
3	10.4	9.5	13.2
4	10.3	9.3	9.8
5	10.1	8.2	12.8
6	9.9	7.1	7.4
7	8.3	6.8	6.1
8	6.3	6.3	8.8
9	7.7	5.3	8.3
10	7.4	2.2	6.6
11	3.2	3.7	
12	0.6	4.5	
13	0.3	3.0	
14		4.8	
15		2.0	

This sample size is also sufficient to test for DIF where, at α of 0.01 a difference of 0.5 SD within the residuals can be detected for any 2 groups with β of 0.20. Bonferroni corrections are applied to both fit and DIF statistics due to the number of tests undertaken (38). A value of 0.005 is used throughout. Due to the floor and ceiling effects observed in the RMA scores of some patients, in order to increase the power of analysis, both admission and discharge data of RMA were pooled for Mokken and Rasch analysis ($n = 209$). A requirement for this is that the scales are invariant across time, and thus DIF by time was examined separately.

Statistical analysis was undertaken with SPSS for Windows 15.0 (Chicago, IL, USA); Mokken scale analysis was undertaken with procedure "msp" within STATA (39) and Rasch analysis with RUMM2020 (40).

RESULTS

Reliability

Cronbach's alpha and ICC values of Gross Function section (RMA-gf) were 0.93 and 0.88; that of Leg and Trunk section (RMA-lt) 0.88 and 0.84 and of the Arm section (RMA-a) 0.95 and 0.93. The Rasch-based person separation index (PSI) of the RMA-gf, RMA-lt and RMA-a were 0.95, 0.89 and 0.98, respectively. Consequently all scales satisfy reliability levels for individual patient use.

Initial scalability

Mokken scale analysis showed that the RMA-lt and the RMA-a both satisfied non-parametric Guttman Scaling requirements with Loevinger's coefficients of 0.723 and 0.927, respectively. Eleven of the 13 items of the Gross Function section (RMA-gf) satisfied those criteria with a Loevinger's coefficient of 0.944. Two items (12 and 13) formed a separate scale. However, the hierarchical ordering of items within each subscale differed from that given by the developers (Table I). For example, item 8 of the Gross Motor Function scale "walking 10 m indoors without aid", had fewer positive responses than the 2 following items (9 and 10), shifting the ordering from that originally reported. In the leg and trunk scale, item 7 "Standing, tap ground lightly" is much harder for this group of patients, and has a lower frequency of success than subsequent items in the original hierarchy.

Internal construct validity

Data from each subscale were then fitted to the Rasch measurement model. In the RMA-gf section, 3 extreme items (items 13, 12, 1) were excluded from the analysis. Item 2 showed misfit to the model. After removal of this item, a 9-item RMA-gf scale satisfied the model expectations with a mean item fit residual –0.633 (SD 0.773), person fit residual –0.270 (SD 0.240), and χ^2 item-trait interaction 19.5 (df 9), $p = 0.02$ (Table II). The independent *t*-test analysis indicated strict unidimensionality (1.7%: confidence interval (CI) 0–5.6%). Some local dependency of items was observed but creating a testlet (adding items together to make a combined score) showed little improvement in overall fit. DIF by time was found only in item 4 (transfer from wheelchair to chair towards unaffected side). The PSI was high at 0.95 and the hierarchical ordering of the items was consistent with the original design with the exception of item 8, "walking 10 m indoors without aid", which was much harder in this group of patients after stroke (Fig. 1a).

In the RMA-lt section, item 7 showed misfit to the model. After removal of this item, a 9-item RMA-lt scale satisfied the model expectations [Mean item fit residual –0.598 (SD 0.657), person fit residual –0.351 (SD 0.560), and χ^2 item-trait interaction 15.4 (df 18), $p = 0.64$] (Table III). The scale was unidimensional (5.8%: CI 2.2–9.4%). Two items demonstrated

Table II. *Fit of Rivermead Motor Assessment – gross function section to the Rasch model (n = 118)*

Item number	Item	Location	SE	Residual	χ^2	df	p
gf3	Sitting to standing	–3.48	0.48	–0.50	0.92	1	0.339
gf4	Wheelchair to chair transfer – unaffected side	–3.08	0.44	–0.71	0.74	1	0.389
gf5	Wheelchair to chair transfer – affected side	–2.61	0.41	–0.89	1.11	1	0.292
gf6	Walk 10 m indoors with aid	–2.41	0.40	–0.60	0.46	1	0.497
gf7	Climb stairs independently	0.28	0.31	–1.22	2.55	1	0.111
gf8	Walk 10 m indoors without aid	3.35	0.26	–0.54	6.65	1	0.010
gf9	Walk – pick up bag from floor	1.20	0.30	–2.01	0.27	1	0.604
gf10	Walk outside 40 m	1.49	0.29	0.83	1.25	1	0.263
gf11	Walk up and down steps	5.27	0.29	–0.05	5.60	1	0.018

SE: standard error; df: degrees of freedom.

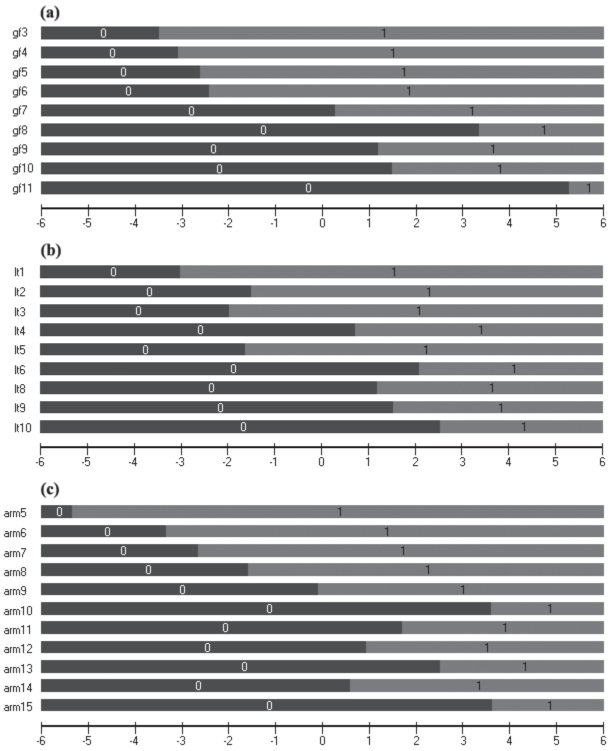

Fig. 1. Hierarchical ordering of items in 3 sections of Rivermead Motor Assessment (RMA). (a) gross function section (gf); (b) leg and trunk section (lt); (c) arm section (arm). 0 = fail; 1 = pass.

local dependency, but combining them made the fit to model expectations considerably worse. Reliability (PSI) was found to be 0.89, but items 4, 8, and 9 showed DIF by age and item 5 by gender. No items showed DIF by time. Hierarchical ordering of the items differed from that originally published (Fig. 1b).

In the RMA-a section, 4 items (items 1, 2, 3, 4) did not fit model expectations. After deletion of these items, an 11-item RMA-a scale satisfied model expectations [Mean item fit residual –0.260 (SD 0.301), person fit residual –0.444 (SD 0.941), and a non-significant χ^2 item-trait interaction 21.2 (df 22), $p = 0.51$] (Table IV). The number of significant independent *t*-tests was low (2.6%: CI 0–7.5%) supporting the unidimensionality of the scale. Some local dependency was observed, but creating a testlet showed little improvement in overall fit. The PSI was high at 0.98, but items 8, 10, and 15 showed DIF by gender and 15 by lesion. Again, no items showed DIF by time. However, the original hierarchical ordering of the scale was not supported with, for example, the item "pat large ball upon floor" being much more difficult than suggested by the original hierarchical ordering (Fig. 1c)

Targeting of the patients in relation to the items for each subsection of RMA is also displayed in Fig. 2. The scatter plot of Rasch transformed measures vs raw scores and 95% confidence limits (the classic Rasch "ogive") is shown in Fig. 3.

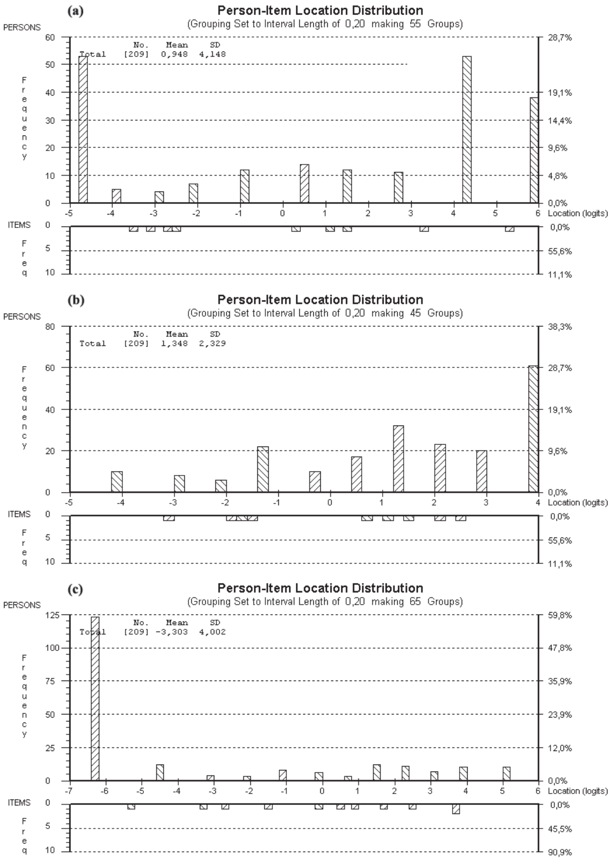

Fig. 2. Targeting of scales to patient ability for: (a) gross function section; (b) leg and trunk section; and (c) and arm section. SD: standard deviation.

Table III. *Fit of Rivermead Motor Assessment – leg and trunk section to the Rasch model (n = 138)*

Item number	Item	Location	SE	Residual	χ^2	df	*p*
lt1	Roll to affected side	–3.01	0.36	–0.53	2.14	2	0.342
lt2	Roll to unaffected side	–1.50	0.27	–0.97	1.52	2	0.469
lt3	Half-bridging	–1.96	0.29	–1.74	1.61	2	0.447
lt4	Sitting to standing	0.72	0.21	–0.38	2.24	2	0.327
lt5	Half-crook lying	–1.63	0.27	–0.68	1.23	2	0.542
lt6	Step unaffected leg	2.10	0.22	–1.12	3.93	2	0.141
lt8	Lying ankle dorsiflexion – leg flexed	1.19	0.21	0.06	0.66	2	0.719
lt9	Lying ankle dorsiflexion – leg extended	1.55	0.21	0.52	0.66	2	0.719
lt10	Stand with affected hip	2.53	0.23	–0.53	1.36	2	0.506

SE: standard error; df: degrees of freedom.

Table IV. *Fit of Rivermead Motor Assessment-arm section to the Rasch model (n = 76)*

Item number	Item	Location	SE	Residual	χ^2	df	*p*
Arm 5	Pick up large ball with 2 hands	–5.35	0.56	–0.23	1.66	2	0.435
Arm 6	Pick up tennis ball from table	–3.33	0.46	–0.88	0.54	2	0.765
Arm 7	Pick up pencil from table	–2.65	0.45	–1.31	1.02	2	0.602
Arm 8	Pick up paper from table	–1.58	0.42	–1.75	3.21	2	0.201
Arm 9	Cut putty with a knife and fork	–0.09	0.37	–0.34	2.12	2	0.346
Arm 10	Pat large ball on floor	3.61	0.37	–0.19	0.25	2	0.881
Arm 11	Fast opposition of thumb	1.71	0.33	1.90	3.97	2	0.137
Arm 12	Fast supination and pronation	0.94	0.34	–0.42	1.06	2	0.590
Arm 13	Horizontal arm abduction – body rotation	2.53	0.33	–0.36	0.29	2	0.863
Arm 14	String around head and tie bow	0.59	0.35	–1.16	5.30	2	0.071
Arm 15	“Pat-a-cake” on wall	3.63	0.37	–0.15	1.81	2	0.404

SE: standard error; df: degrees of freedom.

External construct validity

The correlations of RMA-gf and RMA-lt with FIM™ were moderate to high, whilst RMA-a showed moderate correlations with FIM™ subscales (Table V).

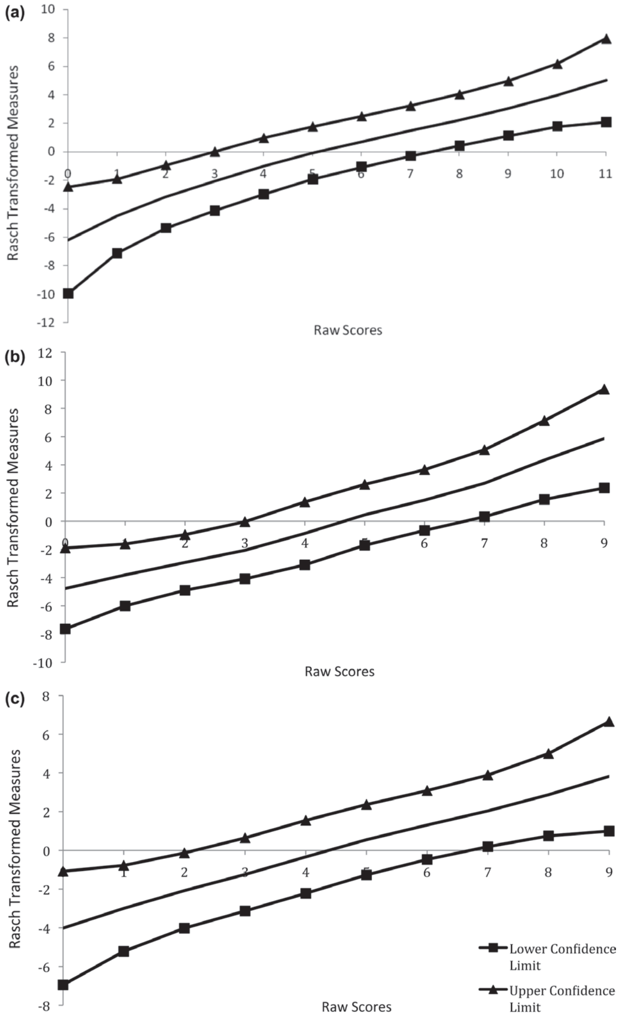

Fig. 3. Scatter plot of Rasch transformed measures vs raw scores and 95% confidence limits. (a) gross function section; (b) leg and trunk section; (c) arm section.

Responsiveness

The ES and SRM were 0.51 and 0.83 for RMA-gf; 0.45, 0.86 for RMA-lt, and 0.38, 0.60 for RMA-a, respectively, whereas the ES and SRM were 0.61 and 1.20 for FIM motor scale.

DISCUSSION

In this study the psychometric properties of all 3 sections of RMA were assessed for scaling properties and internal construct validity by Mokken scaling and Rasch analysis, as well as traditional methods to assess reliability, external construct validity, and responsiveness. The reliability of all 3 sections of the RMA was good when evaluated by both conventional and modern statistical methods. In the previous report by Van de Winckel et al. (16) the PSI of arm section was found to be 0.93, which is in concordance with the findings of the present study.

Mokken scaling was satisfactory, suggesting a probabilistic Guttman hierarchy among the items, but highlighted that this ordering failed to support that originally published. Consequently, although a non-parametric probabilistic Guttman scaling is evident, this will not support the “3 failures” approach used with the originally specified hierarchical ordering.

Internal construct validity results indicated that all 3 sections of the scale required some modification to satisfy Rasch model expectations. This was due to both structural problems with the sample distribution, and to the extra demands placed upon the pattern of item responses required to effect an interval scale transformation. For example, in the RMA-gf section, item 1 (sit

Table V. *Correlations of Rivermead Motor Assessment sections with Functional Independence Measure (FIM™)*

	Admission			Discharge		
	FIM Motor	FIM Self-Care	FIM Mobility	FIM Motor	FIM Self-Care	FIM Mobility
RMA-gf	0.865	0.815	0.844	0.820	0.757	0.817
RMA-lt	0.784	0.726	0.782	0.747	0.702	0.764
RMA-a	0.386	0.390	0.386	0.467	0.480	0.483

Spearman r, $p < 0.001$.
a: arm section; gf: gross function section; lt: leg and trunk section.

unsupported) was achieved by everyone, while items 12 (run 10 m), and 13 (hop on affected foot 5 times) could not be done by anyone. These results are clinically consistent considering the properties of the patient group. Most of the patients were in the subacute phase, so they already managed to sit unsupported. On the other hand, items 12 and 13 are much more difficult than any patient at this stage of stroke recovery could perform. Similar results were observed by Adams et al. (14) in their hierarchical scalability study. Removal of the extreme items and the misfit item (item 2 lying to sitting) satisfied Rasch model expectations. Likewise, in the RMA-gf section, item 7 (standing, tap ground lightly 5 times with unaffected foot) required removal in order to satisfy expectations.

In the RMA-a section, items 1 (lying, protract shoulder), 2 (lying, hold extended arm in elevation), 3 (lying, flexion and extension of elbow), and 4 (sitting, elbow pronation - supination) showed misfit. After these 4 items were excluded, the arm section satisfied the model expectations and was found to be unidimensional. This is contrast to the findings in the study by Van de Winckel et al. (16) where, except for item 1, different items were discarded using the same analytical approach (items 1, 6, 7, 8). This difference may indicate variability of scaling characteristics by diagnostic subgroup or, for example, by time since stroke, which was longer in the study by Van de Winckel et al. (16). Interestingly, the arm section of the RMA showed discrepant scalability characteristics in distinct (acute and non-acute) stroke populations in earlier studies (13, 14). These results give rise to an uncertainty about the stability of the arm section of the scale. In the present study, items of all 3 sections were mostly free of clinically important DIF, which is another component of internal construct validity.

The original RMA was assumed to be a hierarchical scale. Thereby, in each section, the evaluation could be stopped after 3 consecutive item failures. Later on, revised guidelines recommended that all the items of leg and trunk section should be performed, even in cases of 3 consecutive failures. However, one of the main results of the present study has revealed that the hierarchical ordering of the items in all 3 sections differed from that originally specified. Similar findings were reported for the arm section in the study by Van de Winckel et al. (16), and previous work has shown that arm and leg and trunk sections did not meet the hierarchical scaling criteria using Guttman technique in different stroke populations (13, 14). Thus, the current study supports the recommendation that all items be scored on each scale, in both clinical assessments and trials. Nevertheless, the presence of extreme items suggests that such an approach (i.e. stop after 3 failures) would be useful to avoid over-burdening patients with too-difficult (or too-easy) tasks. Consequently the challenge is to identify how the hierarchical ordering varies by diagnostic subgroup and in different settings such that a more accurate tailored "stop" routine may be employed.

To our knowledge, the sensitivity to change of the RMA has not been investigated extensively previously. In only 2 studies the authors implied that the RMA was sensitive to change, but in both studies there were no quantitative evaluations (12, 13). In the present study, ES and SRM values showed that all 3 sections of the RMA were sensitive to change, the arm section being the least responsive when compared with the FIM™ motor scale. Expected moderate, and moderate to high correlations observed between the RMA sections and the FIM™ motor scale supported the external construct validity of the scale.

There are some weaknesses of the study. The sample size is low and may not be heterogeneous enough regarding time since stroke, as 70% of the group was at 1–3 months after onset of stroke. Another problem is the number of extreme cases. Of the 214 assessments, 43.5% of those for the gross function section, 34% of the leg and trunk section, and 63.6% in the arm section were excluded from the Rasch analysis due to both ceiling or floor effects, and thus a limited number of cases remained for the final analysis. This is, of course, not a problem only for Rasch analysis, as substantial floor and ceiling effects compromise the validity of the scale in any setting. Nevertheless, the analysis of the scales was less than optimal.

In conclusion, this paper has evaluated the 3 sections of the RMA by modern psychometric approaches and has shown that, while a Guttman ordering is evident, the hierarchy of items differs from that published, and that as a consequence the "stop after 3 failures" approach using the original hierarchical ordering will be compromised. The reliability of the scale has been shown by both internal consistency and ICC, indicating a level consistent with individual use. The responsiveness has also been confirmed quantitatively by both ES and SRM. Furthermore, with some modifications, the data satisfy the most stringent measurement criteria defined by modern psychometric theory and consequently can provide interval scale estimates of motor function in stroke. Some of these modifications were necessitated by extreme items, and may not be required in a sample with much broader levels of activity limitations.

Consequently, these results need to be verified in larger and heterogeneous stroke populations where floor and ceiling effects are minimized. For the present, until further work demonstrates the robustness of the above findings, and where clinical management need may require information from all items, the original item set can be used. However, we recommend scoring the scales in the original form (for comparative purposes) and in their revised form (excluding deleted items) for statistical purposes (where the summed score is valid and unidimensional) until the stability of the item set has been confirmed. Where change scores are required, or calculations of responsiveness are to be made, these require Rasch-transformed interval-scale scores.

REFERENCES

1. Harvey RL. Improving poststroke recovery: neuroplasticity and task-oriented training. Curr Treat Options Cardiovasc Med 2009; 11: 251–259.
2. Machado S, Bittencourt J, Minc D, Portella CE, Velasques B, Cunha M, et al. Therapeutic applications of repetitive transcranial magnetic stimulation in clinical neurorehabilitation. Funct Neurol 2008; 23: 113–122.

3. Lincoln N, Leadbitter D. Assessment of motor function in stroke patients. Physiotherapy 1979; 65: 48–51.
4. De Wit L, Putman K, Schuback B, Komárek A, Angst F, Baert I, et al. Motor and functional recovery after stroke: a comparison of 4 European rehabilitation centers. Stroke 2007; 38: 2101–2107.
5. Mayr A, Kofler M, Quirbach E, Matzak H, Fröhlich K, Saltuari L. Prospective, blinded, randomized crossover study of gait rehabilitation in stroke patients using the Lokomat gait orthosis. Neurorehabil Neural Repair 2007; 21: 307–314.
6. Woldag H, Gerhold LL, de Groot M, Wohlfart K, Wagner A, Hummelsheim H. Early prediction of functional outcome after stroke. Brain Inj 2006; 20: 1047–1052.
7. Johnson CA, Burridge JH, Strike PW, Wood DE, Swain ID. The effect of combined use of botulinum toxin type A and functional electric stimulation in the treatment of spastic drop foot after stroke: a preliminary investigation. Arch Phys Med Rehabil 2004; 85: 902–909.
8. van Vliet PM, Lincoln NB, Foxall A. Comparison of Bobath based and movement science based treatment for stroke: a randomised controlled trial. J Neurol Neurosurg Psychiatry 2005; 76: 503–508.
9. Werner C, Von Frankenberg S, Treig T, Konrad M, Hesse S. Treadmill training with partial body weight support and an electromechanical gait trainer for restoration of gait in subacute stroke patients: a randomized crossover study. Stroke 2002; 33: 2895–2901.
10. Lincoln NB, Parry RH, Vass CD. Randomized, controlled trial to evaluate increased intensity of physiotherapy treatment of arm function after stroke. Stroke 1999; 30: 573–579.
11. Collen FM, Wade DT, Bradshaw CM. Mobility after stroke: reliability of measures of impairment and disability. Int Disabil Studies 1990; 12: 6–9.
12. Collin C, Wade D. Assessing motor impairment after stroke: a pilot reliability study. J Neurol Neurosurg Psychiatry 1990; 53: 576–579.
13. Adams SA, Ashburn A, Pickering RM, Taylor D. The scalability of the Rivermead Motor Assessment in acute stroke patients. Clin Rehabil 1997; 11: 42–51.
14. Adams SA, Pickering RM, Ashburn A, Lincoln NB. The scalability of the Rivermead Motor Assessment in nonacute stroke patients. Clin Rehabil 1997; 11: 52–59.
15. Tesio L, Simone A, Bernardinello M. Rehabilitation and outcome measurement: where is Rasch analysis-going? Eura Medicophys 2007; 43: 417–426.
16. Van de Winckel A, Feys H, Lincoln N, de Weerdt W. Assessment of arm function in stroke patients: Rivermead Motor Assessment arm section revised with Rasch analysis. Clin Rehab 2007; 21: 471–479.
17. Mokken RJ. The theory and procedure of scale analysis with applications in political research. New York: Walter de Gruyter, Mouton; 1971.
18. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. J Clin Epidemiol 1988; 41: 105–114.
19. Hamilton BB, Granger CV, Sherwin FS, Zielezny M, Tashman JS. Uniform national data system for medical rehabilitation. In: Fuhrer MJ, editor. Rehabilitation outcomes: analysis and measurement. Baltimore, MD: Paul H. Brookes Publishing Co.; 1987, p. 137–147.
20. Küçükdeveci AA, Yavuzer G, Elhan AH, Sonel B, Tennant A. Adaptation of the Functional Independence Measure for use in Turkey. Clin Rehabil 2001; 15: 311–319.
21. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000; 25: 3186–3191.
22. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951; 16: 297–334.
23. Shrout PE, Fleiss JL. Intra class correlations: uses in assessing rater reliability. Psychological Bull 1979; 86: 420–428.
24. Streiner DL, Norman GR. Health measurement scales. Oxford: Oxford University Press; 1995.
25. Mokken RJ, Lewis C. A nonparametric approach to the analysis of dichotomous item responses. Appl Psychol Meas 1982; 6: 417–430.
26. Molenaar IW. A weighted Loevinger H-coefficient extending Mokken scaling to multicategory items. Kwantitatieve Methoden 1988; 9: 115–126.
27. van Shuur WH. Mokken scale analysis: between the Guttman scale and parametric item response theory. Polit Anal 2003; 11: 139–163.
28. Christensen KB , Kreiner S. Monte Carlo tests of the Rasch model based on scalability coefficients. Br J Math Stat Psychol 2009 Apr 1 [Epub ahead of print].
29. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1960 (reprinted 1980).
30. Andrich D. Rasch models for measurement. London: Sage Publications; 1988.
31. Fischer GH, Derivations of the Rasch model. In: Fischer GH, Molenaar IW, editors. Rasch models; foundations, recent developments and applications. Springer-Verlag: New York; 1995.
32. Pallant JF. Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Brit J Clin Psychology 2007; 46: 1–18.
33. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Care Res 2007; 57: 1358–1362.
34. Angoff WH. Perspectives on differential item functioning methodology. In: Holland PW, Wainer H, editors. Differential item functioning. Hillsdale, New Jersey: Lawrence Erlbaum; 1993, p. 3–23.
35. Smith EV. Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Measure 2002; 3: 205–231.
36. Wright JG, Young NL. A comparison of different indices of responsiveness. J Clin Epidemiol 1997; 50: 239–246.
37. Linacre JM. Sample size and item calibration stability. Rasch Measure Trans 1994; 7: 28.
38. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995; 310: 170.
39. StataCorp. Stata Statistical Software: Release 10. College Station, TX: StataCorp LP; 2007.
40. Andrich D, Lyne A, Sheridon B, Luo G. RUMM 2020. Perth: RUMM Laboratory; 2003.