

ORIGINAL REPORT

STROKE SELF-EFFICACY QUESTIONNAIRE: A RASCH-REFINED MEASURE OF CONFIDENCE POST STROKE

Afsane Riazi, PhD¹, Trefor Aspden, PhD¹ and Fiona Jones, PhD²

From the ¹Department of Psychology, Royal Holloway, University of London, Surrey and ²Faculty of Health, Social Care and Education, Kingston University and St. George's University of London, London, UK

Objective: Measuring self-efficacy during rehabilitation provides an important insight into understanding recovery post stroke. A Rasch analysis of the Stroke Self-efficacy Questionnaire (SSEQ) was undertaken to establish its use as a clinically meaningful and scientifically rigorous measure.

Methods: One hundred and eighteen stroke patients completed the SSEQ with the help of an interviewer. Participants were recruited from local acute stroke units and community stroke rehabilitation teams. Data were analysed with confirmatory factor analysis conducted using AMOS and Rasch analysis conducted using RUMM2030 software.

Results: Confirmatory factor analysis and Rasch analyses demonstrated the presence of two separate scales that measure stroke survivors' self-efficacy with: i) self-management and ii) functional activities. Guided by Rasch analyses, the response categories of these two scales were collapsed from an 11-point to a 4-point scale. Modified scales met the expectations of the Rasch model. Items satisfied the Rasch requirements (overall and individual item fit, local response independence, differential item functioning, unidimensionality). Furthermore, the two subscales showed evidence of good construct validity.

Conclusions: The new SSEQ has good psychometric properties and is a clinically useful assessment of self-efficacy after stroke. The scale measures stroke survivors' self-efficacy with self-management and activities as two unidimensional constructs. It is recommended for use in clinical and research interventions, and in evaluating stroke self-management interventions.

Key words: stroke; self-efficacy; Rasch; measurement; questionnaire; scale.

J Rehabil Med 2014; 46: 406–412

Correspondence address: Dr. Afsane Riazi, Department of Psychology, Royal Holloway, University of London, Surrey, UK TW20 0EX. E-mail: Afsane.Riazi@rhul.ac.uk

Accepted Dec 10, 2013; Epub ahead of print Mar 24, 2014

INTRODUCTION

Stroke survivors with disabilities need to make adjustments to their lives, often by learning new behaviours and modifying lifestyles. The acquisition of such skills is contingent on factors such as the quality and intensity of rehabilitation (1) and a person's psychological state (2).

Self-efficacy is a psychological construct defined as 'the beliefs in one's capabilities to organise and execute the courses of action required to produce attainments' (3). Self-efficacy plays a role as a cognitive mediator of action (4). It has been demonstrated to mediate the influence of a wide range of external variables (e.g. psychological interventions) on behaviour, and higher levels of self-efficacy are associated with higher performance attainment (4). Self-efficacy predicts quality of life, mood and functional independence post stroke (5). Greater awareness of an individual's self-efficacy during rehabilitation could provide better understanding of both successes and lack of progress in attainment of function and accomplishment of goals. By accurately measuring self-efficacy, any mismatches between anticipated and ultimate level of function can be highlighted. This knowledge can benefit clinical practice and the design and development of clinical trials in stroke rehabilitation.

There is increasing interest in supporting stroke survivors and their carers to self-manage the long-term effects of stroke. Because self-efficacy influences initiation of behaviour change (3), it is a cornerstone of many self-management programmes (SMPs). Stroke-specific SMPs have demonstrated benefits (6). However the optimal timing and format for delivering such interventions are still unclear (5). In order to be maximally effective, SMPs need to be designed to enhance self-efficacy (7). A valid measure of self-efficacy will allow for the identification of the components of SMPs that enhances self-efficacy, leading to their improved design (8).

The Stroke Self-Efficacy Questionnaire (SSEQ) is a self-report scale of self-efficacy post stroke that demonstrates good internal consistency and criterion validity (9). Although Principal Components Analysis has provided evidence for a single-factor solution, theoretically, SSEQ may measure two separate dimensions of self-efficacy relevant to recovery and independence after stroke (activity and self-management). These two aspects of self-efficacy are seemingly conceptually distinct. Self-efficacy in functional activity is not necessarily linked to self-efficacy in self-management after stroke. There is evidence that self-efficacy in one domain is not necessarily linked to self-efficacy in another (4). For example, in the exercise behaviour domain task efficacy, coping efficacy, and scheduling efficacy have been demonstrated to be separate dimensions of exercise self-efficacy, with changes in each of these 3 dimensions being relatively independent of changes

in the other dimensions (10). This suggests that these two dimensions need to be assessed separately.

Further evaluation of the SSEQ using modern psychometric methods is needed. The Rasch measurement model is the most commonly used of these newer approaches (11). It examines wider attributes not analysed by classical test theories, e.g. how well an item performs in terms of its relevance or usefulness for measuring the underlying construct, the possible redundancy of an item, and the appropriateness of the response categories (12–15). In particular, the latter refers to whether individuals can be differentiated by their responses as clearly as the categories allow (16), as a higher number of scoring points may not lead to finer differentiation of the participants (17).

Aims

Several outcome measures of functioning post stroke have either been Rasch-built (18, 19) or Rasch-refined (20). The objectives of this study were to perform a Rasch analysis of the SSEQ, to investigate the factor structure of the SSEQ, and to revise the SSEQ so that it meets the current standards of a rigorous and clinically useful rating scale.

METHODS

Participants and settings

Stroke survivors between 4–24 weeks after their first stroke were recruited from local acute stroke units and community stroke rehabilitation teams. All participants were over 18, had a diagnosis of stroke confirmed by scan, and provided full informed consent. Participants were excluded if they were unable to read or had difficulties understanding the questionnaire instructions. Measures were administered by an interviewer either in hospital or home setting. London Surrey Borders ethics committee approved the study.

Measures

The SSEQ is a 13-item self-report scale measuring self-efficacy judgements in specific domains of functioning post stroke. Individuals rate their belief in their ability to achieve each of the 13 items on a 10-point scale, where 0 = not at all confident to 10 = very confident. Preliminary psychometric evaluation has been reported (9).

For validity testing, 4 other measures were administered: the Falls Efficacy Scale (FES) (21), the Modified Rivermead Mobility Index (MRMI) (22), the Hospital Anxiety and Depression Scale (HADS) (23), and the Rosenberg Self-esteem Scale (24). The 10-item FES demonstrates reliability and validity and specifically measures confidence and fear of falling. The MRMI is a 15-item measure of functional mobility following stroke with acceptable levels of reliability, validity and responsiveness in a sample of stroke survivors (25). The HADS is a 14-item scale measuring anxiety and depression precluding somatic symptoms of illness (26). The Rosenberg Self-esteem Scale is a 10-item scale of global self-esteem measuring overall feelings of self-worth (e.g. 27).

Analyses

Confirmatory Factor Analysis. Confirmatory factor analysis (CFA) conducted in AMOS 16 was used to establish whether, as hypothesised, the SSEQ consists of the two unidimensional subscales: Activity (items 1 to 8) and Self-management (items 9 to 13). A single factor model consisting of a single latent variable, loading onto all 13 items of the SSEQ, was compared to a two factor model consisting of 2 correlated latent variables (one of which loaded onto items 1 through 8, the

other loading onto items 9 through 13). To determine which model fit the data best, model fit, standardised residual covariances (SRCs), and factor loadings were examined. Model fit was determined using Comparative Fit Index (CFI) and Standardised Root Mean Square Residual (SRMR), values of CFI ≥ 0.95 and SRMR ≤ 0.09 being taken to indicate good model fit (28). SRCs were examined for values < -2.0 and > 2.0 , which can indicate that items share less or more variance than suggested by the model, respectively. Factor loadings were examined for loadings < 0.40 .

Rasch analyses. Rasch analysis is a probabilistic mathematic modelling technique used to assess properties of outcome measures (29). Where data are shown to accord with model expectations, the internal construct validity of the scale is supported, and ordinal data is transformed to interval scaling (29, 30). During this process, key attributes, such as unidimensionality, invariance, sample targeting of a scale, the appropriateness of response options, hierarchy of item difficulty and local independence of items are examined (31). Further details regarding Rasch analysis is discussed elsewhere (e.g. 14). The analysis was conducted in RUMM2030 (32). As the likelihood ratio test was significant for both scales ($p < 0.05$), both scales were analysed using the unrestricted partial-credit model. Specifically, we looked at the following 7 criteria (29):

1. *Item response-threshold ordering.* Each item had 11 response categories (levels), resulting in 10 item-response-thresholds of increasing order. If categories are discriminating correctly, subjects with higher ability will score higher than subjects with lower ability, indicating correct category discrimination. However, response categories do not always correctly discriminate between subjects of different ability levels, and do not provide additional information about a person's true location on the variable. In such instances, Rasch analysis can be used to reduce the number of response categories by combining two or more categories into a single category.

2. *Tests of fit to the model.* The extent of the fit between the observed data and the Rasch model was examined at both scale and item levels using a combination of fit residuals, chi-square statistic, and item characteristic curves (intraclass correlation coefficients; ICCs). Fit residuals represent the difference between observed and expected responses to an item across participants. At the item level these should fall between ± 2.5 (positive residuals indicate misfit to the Rasch model; negative residuals indicate an item shares a large portion of variance with other items and so is redundant). Chi-square values summarize the difference between observed and expected responses to an item for class intervals of people with similar levels of disability, and should be non-significant ($p > 0.05$, Bonferonni adjusted). Summary chi-square for the scale should also be non-significant ($p > 0.05$). ICCs provide a graphical representation of observed and expected responses for different class intervals of people, allowing a visual examination of fit.

3. *Reliability.* Person Separation Index (PSI) indicates the extent to which the questionnaire distinguishes distinct ability levels. This is equivalent to Cronbach's alpha, with values ≥ 0.80 regarded as adequate.

4. *Item difficulty and person ability.* The location on the Rasch continuum, measured in logits. Ideally, items will have locations that are evenly spaced over a wide range rather than being clustered at the same location. Also, ideally this range of item locations will adequately target the range of locations of the target population.

5. *Local dependency.* High standardised residual correlations (SRCs) between two items indicate that the responses covary greater than predicted by the Rasch model. This indicates the presence of local dependency, where the response to one item is dependent on the response to another item. SRCs > 0.3 can indicate local dependency among items.

6. *Differential item functioning (DIF).* We examined the extent to which items functioned differently by gender (50.0% male, 50.0% female) and marital status (46.8% living alone, 53.2% cohabiting). Items were deemed to function differently if DIF was significant at the $p < 0.05$ (Bonferonni adjusted) level.

7. *Unidimensionality.* An assumption of the Rasch model is that the items of a scale form a unidimensional variable. Unidimensionality of

scales was tested using independent *t*-tests to compare person scores from the two most divergent subsets of items (33). For this procedure, subsets were formed by selecting the items loading most positively and most negatively on the first factor extracted in a principle component analysis of residuals. If no more than 5% of *t*-tests are significant, or if more than 5% are significant but the lower confidence interval corresponds to less than 5%, the assumption of unidimensionality has not been violated.

Analysis was conducted on a sample of 118 participants.

Scale validity. Convergent and discriminant construct validity was determined by the strength of Pearson's correlations between SSEQ scales and other measures. The predicted strengths of the correlations were based on the similarity/dissimilarity of constructs being measured by the scales. For the Activity scale, highest correlations with FES and MRMI, moderate correlations with HADS-depression and HADS-anxiety, and lowest correlations with the Rosenberg Self-esteem Scale (as the Activity scale relates to functional mobility) were expected. For the self-management scale, highest correlations with the HADS-depression and HADS-anxiety and the Rosenberg Self-esteem Scale, and lowest correlations with FES and MRMI were expected.

RESULTS

Demographic details

A total of 118 stroke survivors took part in the study (85% response rate). Fifty-four percent were recruited from the community, 27% from a rehabilitation centre and 18% from the acute ward. The mean days since stroke was 122.3 (standard deviation; SD 94.7), 50% were female, the mean age was 68.4 years (SD 16.1), 46% of participants were living alone and 81% were white.

Confirmatory factor analysis

The single factor model showed a poor fit to the data; $\chi^2(65)=387.81$, CFI=0.69, SRMR=0.12. Additionally, several high (>2.0) SRCs amongst items 9 through 13, as well as negative (<-2.0) SRCs between items 1 through 8 and items 9 through 13, suggested that items 9 through 13 formed a separate scale to items 1 through 8. For the two-factor model, whilst not showing a particularly good fit to the data, the fit to the data was considerably better than for the single-factor model; $\chi^2(64)=247.82$, CFI=0.82, SRMR=0.10, with a correlation between the two factors of 0.58, suggesting that a two factor model is more appropriate for these items than a single-factor model. Based on this analysis, further analyses were conducted separately on an Activity scale (items 1 through 8) and a Self-management scale (items 9 through 13).

Rasch analysis

Appropriateness of the number of response options. The majority of items across both the Activity and Self-management scales were shown to have disordered response thresholds (e.g. see category probability curve in Fig. 1). The disordering was most apparent between the lower categories 0, 1, 2, and 3, as well as between the higher categories 8, 9, and 10. Examination of category frequencies showed several of these lower and higher categories were sparsely populated. A number of rescoring options were tested. Rescoring all items from the 11-point scale to a 4-point scale resulted in the most

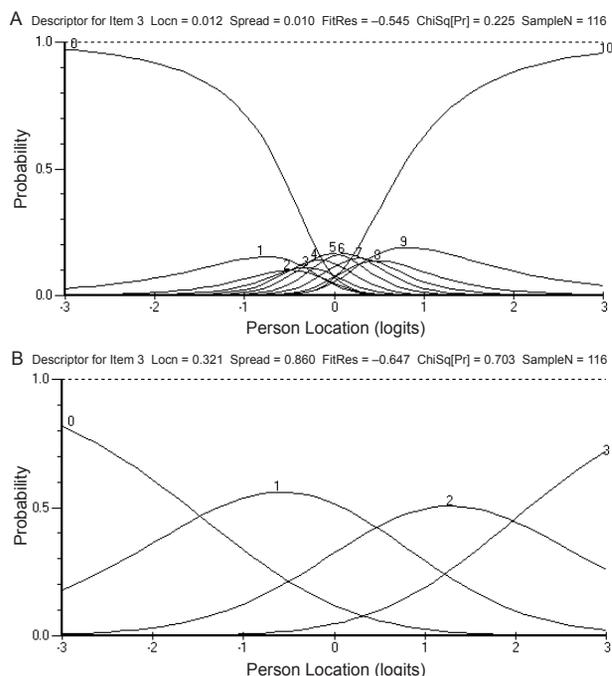


Fig. 1. Category probability curves for item 3 (walk inside) in the Activity scale before (A) and after (B) item rescoring.

appropriate solution for both scales. Specifically, the lowest scoring option (0) was kept as the lowest response option (not at all confident), items 1 through 5 were rescored as option 1 (corresponding to some confidence), items 6 through 9 were rescored as option 2 (corresponding to moderate confidence), and the highest option (10) was kept (very confident). With this solution, all items had ordered response thresholds.

Model fit and reliability

Rescoring items had the effect of improving model fit as represented by the item-trait interaction chi-square statistics: the chi-square for each scale was reduced, with a corresponding increase in associated *p*-value. After rescoring items, both the Activity and Self-management scales demonstrated good fit to the Rasch model (Table I). Summary chi-square for both scales had non-significant *p*-values, and reliability was good (PSI ≥ 0.80 for both scales). The refined scales also demonstrated satisfactory fit residuals for both items (Activity scale mean -0.10 (SD 1.28); Self-management scale mean -0.06 (SD 0.75)) and persons (Activity scale mean -0.46 (SD 1.40); Self-management scale mean -0.76 (SD 1.46)).

At the individual item level, all items fit within the expectations of the Rasch model as demonstrated by fit residuals falling between ± 2.5 and non-significant item chi-square values (Table II). Graphical examination of fit using ICCs also showed close fit to the Rasch model for all items.

Item difficulty and person ability

The calibration of the refined Activity and Self-management scales are displayed in Table II. Items in each scale are ordered according to their ability level. Targeting of the Activity and

Table I. Summary fit statistics after adjusting for local dependency and DIF

Scale	Item fit residual Mean (SD)	Person fit residual Mean (SD)	Person location		Item-trait interaction		
			Range	Mean (SD)	Chi-square (df)	<i>p</i>	PSI
Activity scale	0.00 (0.92)	-0.39 (1.26)	-5.23 to 4.80	0.60 (1.78)	19.93 (16)	0.223	0.87
Self-management scale	-0.08 (0.78)	-0.67 (1.32)	-5.25 to 5.06	1.02 (1.93)	20.50 (12)	0.058	0.80

SD: standard deviation; df: degrees of freedom; PSI: Person Separation Index.

Self-Management scales are displayed in Figs 2a and 2b. The top panel displays the distribution of participants on each scale. For the Activity scale, no participant answered all items across the scale as not at all confident (0 % floor effect) and only two participants answered all items as very confident (1.7% ceiling effect), representing a lack of floor and minimal ceiling effects. For the Self-management scale, only one participant answered all items across the scale as not at all confident (0.8% floor effect) and only 9 participants answered all items as very confident, thus there are minimal floor effects and only a modest (7.6% of participants) ceiling effect.

Local dependency

From the 8-item Activity scale, items 3 (“Walk a few steps on your own on any surface inside your house”) and 4 (“Walk about your house to do most things you want”) demonstrated local dependency (SRC=0.35). To accommodate this local dependency the two items were combined into a testlet, and were treated as a single item for the remainder of this analysis.

Differential item functioning

Item 6 (use both your hands for eating food) from the Activity scale demonstrated DIF by marital status, and Item 10 (do your own exercise program) from the Self-management scale demonstrated DIF by gender. In order to adjust for DIF, Item 6 was split by marital status and Item 10 was split by gender,

with the resultant split items being used in the remainder of the analysis. No items demonstrated DIF for the interaction of any of the person factors with class interval.

Tests of unidimensionality

For the Activity scale, 4.24% of *t*-tests were significant ($p < 0.05$). For the Self-management scale, 2.54% of *t*-tests were significant. Thus for both scales of the SSEQ, the principle of unidimensionality were not violated.

Validity

Convergent and discriminant validity of the refined scales were supported. The direction, magnitude and pattern of correlations were generally consistent with predictions outlined in the Methods section. For the Activity scale, highest correlations were obtained with FES (0.75), moderate correlations with HADS (HADS total -0.56, HADS Anxiety subscale -0.45, HADS Depression subscale -0.59) and MRMI (0.51), and lowest correlations with the Rosenberg Self-esteem Scale (0.32). For the Self-management scale, highest correlations were obtained with HADS (HADS total -0.63, HADS Anxiety subscale -0.53, HADS Depression subscale -0.64), and lower correlations with FES (0.44) and MRMI (0.23). However, lower correlations with the Rosenberg Self-esteem Scale (0.33) were also obtained, confirming that they measure separate constructs.

Table II. Item fit statistics for the Activity and Self-management scales after adjusting for local dependency and DIF

Scale Item	Location (logit)	SE	Threshold 1 logit	Threshold 2 logit	Threshold 3 logit	Item Fit residual	Chi-square	<i>p</i> ^a
<i>Activity</i>								
1: Into bed	-2.992	0.158	-10.17	-0.17	1.37	1.206	0.090	0.956
2: Exit bed	-0.619	0.155	-3.97	0.28	1.84	-0.492	3.306	0.191
7: Dress	-0.021	0.152	-2.54	0.22	2.26	-1.527	7.473	0.024
6: Eat food (live alone)	0.040	0.222	-2.31	0.55	1.88	-0.666	2.220	0.330
8: Prepare meal	0.342	0.146	-1.92	0.71	2.23	0.249	1.092	0.579
3+4 testlet: Walk inside ^b	0.725	0.090	-0.45	-0.57	-0.18	-0.186	0.683	0.711
5: Walk outside	1.239	0.141	-0.54	1.13	3.13	0.275	1.164	0.559
6: Eat food (cohabit)	1.286	0.193	-0.04	0.92	2.98	1.132	3.903	0.142
<i>Self-management</i>								
9: Make progress	-0.704	0.174	-4.17	-0.39	2.45	-0.827	7.905	0.019
10: Own exercise (male)	-0.691	0.230	-3.15	-0.56	1.63	-0.954	2.056	0.358
13: Get faster	0.199	0.164	-2.51	0.16	2.94	0.957	3.895	0.143
10: Own exercise (female)	0.302	0.225	-2.74	1.34	2.31	0.280	1.764	0.414
11: Cope	0.421	0.182	-3.23	0.62	3.87	-0.488	2.610	0.271
12: Do things	0.473	0.156	-2.95	0.95	2.66	0.525	2.269	0.322

^aBonferroni adjusted *p*-values for Chi-square are 0.0063 for the Activity scale and 0.0100 for the Self-management scales (unadjusted $p = 0.05$).

^bThe testlet has 6 thresholds due to combining two items, each with 3 thresholds. Logit locations of the remaining 3 thresholds are: threshold 4=0.63, threshold 5=1.77, threshold 6=3.14.

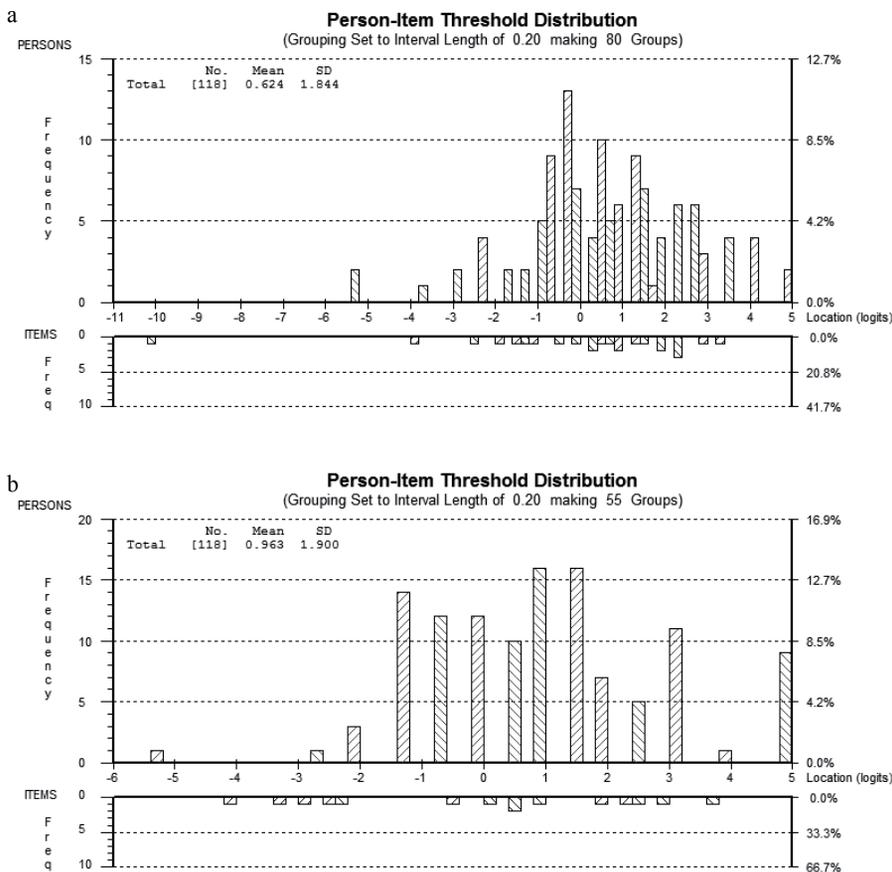


Fig. 2. a) Person-Item Threshold Distributions for Activity scale; b) Person-Item Threshold Distributions for Self-management scale. SD: standard deviation.

Summary

After adjusting for local dependency and DIF, the revised SSEQ satisfies the Rasch requirements. Both Activity (–8.0 to 4.9) and Self–management scales (–5.3 to 5) have wide ranges of scores.

DISCUSSION

The aim of this study was to evaluate the SSEQ as a clinically useful and scientifically rigorous measure of post-stroke self-efficacy. We wished to improve the SSEQ using Rasch analyses and CFA by refining the rating scale and distinguishing items belonging to the same construct. The fact that neither the one-factor or the two-factor models of the SSEQ fit the data well in CFA analysis may be due to local dependency, which had not been adjusted for at this stage of the analysis (see e.g. (34), in which the fit to the data was initially poor but improved once local dependency was adjusted for). Despite this poor fit, the fact that fit was better in the two-factor model suggests a two factor structure to the SSEQ, and results from the Rasch analysis confirm that there are two distinct unidimensional constructs of self-management self-efficacy and activity self-efficacy being measured by the SSEQ.

Our comprehensive evaluation using Rasch analysis demonstrated that, as expected, most original SSEQ items displayed

disordered thresholds, implying that the categories were unable to distinguish the true ability of the respondents as finely as suggested by the 11 categories (35). The modified SSEQ with collapsed categories now adhere to the existing quality criteria for rating scales (15). By collapsing categories with disordered thresholds, redundant categories are eliminated and each category represents a distinct level of ability (36, 37).

Both the modified Self-management and Activity scales demonstrated good model fit. There was no local dependency, the scales have good internal consistencies, and good invariance by group. The two modified scales are unidimensional, and measure two separate constructs. This seems reasonable as stroke survivors may have different levels of self-efficacy for performing activities of daily living as they would for being able to self-manage, as these are conceptually distinct, so someone with high self-efficacy for performing various activities may not have high self-efficacy for self-management. Therefore, obtaining a single total score by summing all items is not recommended, as unidimensionality is a prerequisite to the summation of items (38).

The items provide clinically important information about confidence levels of stroke survivors. Logit ranges were fairly wide for both scales. The targeting and spread of item-difficulty and the high person-separation reliability suggest that the SSEQ is appropriate for measuring with satisfactory

precision persons with a wide variety of self-efficacy beliefs. Ceiling and floor effects were not apparent for the Activity scale, and minimal ceiling effect (<10%) were found for the Self-management scale. The high item-separation reliability indicates that confidence can be placed in the replicability of item-placement across future samples.

As further evidence of construct validity, the hierarchical arrangements of items were consistent with clinical expectations. For example, in the Activity scale, participants rated being the least confident of “walking outside”, and most confident with “getting into bed”. Furthermore, the scales demonstrate good external construct validity as demonstrated by correlations with other widely used measures.

Clinicians can use the SSEQ to gain a greater understanding of individual confidence in the recovery period post stroke and target rehabilitation interventions accordingly. Measuring self-efficacy before and after rehabilitation could enable clinicians to monitor responses to specific interventions. It could be invaluable for clinicians’ greater awareness of those patients that have low or high self-efficacy, and who may need a different approach to their rehabilitation. As such, we recommend that the SSEQ is used alongside other measures of functional performance, e.g. the Barthel Index, to gain greater insight into the relationship between self-efficacy and other stroke outcomes.

The SSEQ can also be used as a secondary outcome measure in both stroke treatment and rehabilitation trials. In rehabilitation studies, as well as demonstrating that the intervention is effective, it is important to understand whether the benefit may generalize beyond the research setting, and whether self-efficacy will play a significant role (39). Thus, in stroke rehabilitation, trials can include primary outcome measure directed at the activity, and could also include self-efficacy as a secondary outcome measure to provide data both on whether the participants can perform the activity, and whether they believe they can perform the activity (39). Self-efficacy is an important secondary outcome measure for evaluation of more complex treatments for similar reasons (39).

There are some limitations with the study. Firstly, although every attempt was made to include people with aphasia, patients who were unable to understand the questionnaire instructions were excluded. This may mean that many patients who did not meet this criteria, but for whom rehabilitation is nonetheless important, were excluded. Similarly, the measure has been developed in the UK sample, and further testing of the measure in a non-UK sample is necessary, as self-efficacy judgements in stroke may be influenced by societal and cultural factors. Further, most of the participants were 3–4 months post stroke, which coincides with getting ready for discharge from rehabilitation, and clinicians need to have an understanding of individual confidence when preparing a person for discharge from rehabilitation. However, future studies could also include patients with a more recent stroke that will undergo rehabilitation strategies. Further studies need to be conducted to identify how these self-efficacy judgements are associated with

biopsychosocial outcomes over time. Future studies should also examine concurrent construct validity with other measures of self-efficacy, especially generic self-efficacy measures, though as far as we are aware, the SSEQ is the only measure of self-efficacy specifically developed for stroke survivors. Lastly, future studies should also examine the performance of the new response categories (to examine whether the scale structures are replicated, as collapsing the categories is not the same as presenting 4 categories to patients), and responsiveness. This should be done ideally in a much larger sample, as we acknowledge that the study may not have been fully powered (for example, to detect misfit or DIF). In the case of DIF, we found that those living alone tended to endorse the item ‘Use both your hands for eating your food’ compared to those living with a spouse or partner, perhaps suggesting that people who live alone develop greater resourcefulness and self-efficacy than those who have support from carers. Further, women found the item ‘Do your own exercise programme every day’ easier to endorse than men of a similar ability level. There is some evidence to suggest that women tend to engage more in self-management programmes, however, a fuller investigation of these differences will be needed in a larger sample.

Measuring and managing self-efficacy through education and rehabilitation programmes has the potential to lead to better outcomes for stroke survivors. A patient may perform well on observed measures of function, but not have confidence in their own capability to persevere and cope with setbacks particularly after being discharged from rehabilitation. Having greater knowledge of self-efficacy, and incorporating strategies to enhance confidence into rehabilitation could lead to better outcomes for stroke survivors. Self-efficacy beliefs are modifiable and can be improved by performance accomplishments, vicarious experience and verbal persuasion (3, 5). Stroke self-management programmes based on self-efficacy principles are being developed and tested (6). If enhancing self-efficacy is the main mechanism by which behaviour change and improved health status can be achieved, self-efficacy measures need to be scientifically rigorous and uncontaminated by other psychological states. By utilising Rasch analyses to refine the scales, the SSEQ now fulfils requirements provided by guideline documents for psychometric standards for rating scales (40), and measures constructs that are conceptually different from anxiety, depression, or self-esteem, as evidenced by our construct validity data.

In conclusion, the Rasch-modified SSEQ presents good psychometric properties and provides a precise assessment of self-efficacy beliefs related to everyday tasks and self-management. It is suitable for use in clinical settings, research, and in the evaluation of self-management interventions.

ACKNOWLEDGEMENTS

We thank the participants who took part in this research, Emma Roffey, Sarah Smith, Jacqueline Steward who assisted with data collection. This work was supported by the SouthWest London Academic Network (SWAN) small grant funding.

REFERENCES

- Langhorne P, Bernhardt J, Kwakkel G. Stroke Rehabilitation. *Lancet* 2011; 377: 1693–1672.
- Kendall E, Catalano T, Kuipers P, Posner N, Buys N, Charker J. Recovery following stroke: the role of self-management education. *Social Science and Medicine* 2007; 64: 735–46.
- Bandura A. The nature and structure of self-efficacy. In: Bandura A, editor. *Self-efficacy: the exercise of control*. New York: W.H Freeman and Company; 2007, p. 36–78.
- Bandura A. Self-efficacy mechanism in human agency. *American Psychologist* 1982; 37: 122–147.
- Jones F, Riazi A. Self-efficacy and self-management after stroke: a systematic review. *Disability & Rehabilitation* 2011; 33: 797–810.
- Cadilhac DA, Hoffman S, Kilkenny M, Lindley R, Lalor E, Osborne RH, et al. A Phase II multicentered, single-blind, randomized, controlled trial of the Stroke Self-management Program. *Stroke* 2011; 42: 1673–1679.
- De Silva D. Evidence: helping people help themselves: a review of the evidence considering whether it is worthwhile to support self-management. London: Health Foundation; 2011.
- Hewlett S, Cockshott Z, Kirwan J, Barrett J, Stamp J, Haslock I. Development and validation of a self-efficacy scale for use in British patients with rheumatoid arthritis (RASE). *Rheumatology* 2001; 40: 1221–1230.
- Jones F, Partridge C, Reid F. The stroke self-efficacy questionnaire: measuring individual confidence in functional performance after stroke. *J Nurs Healthc Chronic Illn* 2008; 17: 244–252.
- Rodgers WM, Wilson PM, Hall CR, Fraser SN, Murray TC. Evidence for a multidimensional self-efficacy for exercise scale. *Res Q Exerc Sport* 2008; 79: 222–234.
- Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: MESA Pr; 1980.
- Franchignoni F, Ferriero G, Giordano A, Sartorio F, Vercelli S, Brigatti, E. Psychometric properties of QuickDASH – A classical test theory and Rasch analysis study. *Manual Therapy* 2011; 16: 177–182.
- Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation. *J Rehabil Med* 2003; 35: 105–115.
- Bond TG, Fox CM. *Applying the Rasch model – Fundamental measurement in the human sciences*. 2nd ed. Mahwah, NY: Lawrence Erlbaum Associates; 2007.
- Wolfe EW, Smith Jr EV. Instrument development tools and activities for measure validation using rasch models: part II – validation activities. *J Appl Meas* 2007, 8: 204–34.
- Linacre, JM. Investigating rating scale category utility. *J Outcome Meas* 1999; 3: 103–122.
- Zhu W, Updyke WF, Lewandowski C. Post-hoc Rasch analysis of optimal categorization of an ordered-response scale. *J Outcome Meas* 1997; 1: 286–304.
- Penta M, Tesio L, Arnould C, Zancan A, Thonnard JL. The ABILHAND questionnaire as a measure of manual ability in chronic stroke patients: Rasch-based validation and relationship to upper limb impairment. *Stroke* 2001; 32: 1627–1634.
- Caty GD, Arnould C, Stoquart GG, Thonnard JL, Lejeune TM. ABILOCO: a Rasch-built 13-item questionnaire to assess locomotion ability in stroke patients. *Arch Phys Med Rehabil* 2008; 89: 284–290.
- Duncan PW, Bode RK, Min Lai S, Perera S; Glycine Antagonist in Neuroprotection Americans Investigators. Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Arch Phys Med Rehabil* 2003; 84: 950–963.
- Tinetti ME, Richman D, Powell L. Falls efficacy as a measure of fear of falling. *J Gerontol* 1990; 45: 239–243.
- Lennon S, Johnson L. The Modified Rivermead Mobility Index: validity and reliability. *Disabil Rehabil* 2000; 22: 833–839.
- Snaith RP. The Hospital Anxiety and Depression Scale. *Health and Quality of Life Outcomes* 2003; 1: 29.
- Rosenberg M. *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press; 1965.
- Hsueh I-P, Chun-Hou W, Ching-Fan S, Hsieh C-L. Comparison of psychometric properties of three mobility measures for patients with stroke. *Stroke* 2003; 34: 1741–1745.
- Sagen U, Vik TG, Moum T, Morland T, Finset A, Dammen T. Screening for anxiety and depression after stroke comparison of the Hospital Anxiety and Depression Scale and the Montgomery and Asberg Depression Rating Scale. *J Psychosom Res* 2009; 67: 325–332.
- Vickery CD, Sepehri A, Evans CC. Self-esteem in an acute stroke rehabilitation sample: a control group comparison. *Clin Rehabil* 2008; 22: 179–187.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struc Equ Modeling* 1999; 6: 1–55.
- Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism* 2007; 57: 1358–1362.
- Perline R, Wright BD, Wainer H. The Rasch model as additive conjoint measurement. *Appl Psych Meas* 1979; 3: 237–256.
- Batcho CS, Tennant A, Thonnard J-L. ACTIVLIM-Stroke: A cross-cultural Rasch-built scale of activity limitations in patients with stroke. *Stroke* 2012; 43: 815–823.
- Andrich D, Lyne A, Sheridan B, Luo G. Rumm 2030. Perth, Western Australia: RUMM Laboratory Ltd; 2010.
- Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principle component analysis of residuals. *J Applied Measurement* 2002; 3: 1–26.
- Lundgren-Nilsson A, Jonsdottir IH, Pallant J, Ahlborg Jr G. Internal construct validity of the Shirom-Melamed Burnout Questionnaire (SMBQ). *BMC Public Health* 2012; 12: 1.
- Franchignoni F, Giordano A, Ferriero G, Orlandini D, Amoresano A, Perucca L. Measuring mobility in people with lower limb amputation: Rasch analysis of the mobility section of the prosthesis evaluation questionnaire. *J Rehabil Med* 2007; 39: 138–144.
- Nilsson AL, Sunnerhagen KS, Grimby G. Scoring alternatives for FIM in neurological disorders applying Rasch analysis. *Acta Neurol Scand* 2005; 111: 264–273.
- Kornetti DL, Fritz SL, Chiu YP, Light KE, Veloza CA. Rating scale analysis of the Berg Balance Scale. *Arch Phys Med Rehabil* 2004; 85: 1128–1135.
- Hsueh IP, Wang WC, Sheu CF, Hsieh CL. Rasch analysis of combining two indices to assess comprehensive ADL function in stroke patients. *Stroke* 2004; 35: 721–726.
- Young CA, Mills RJ, Woolmore J, Hawkins CP, Tennant A. The unidimensional self-efficacy scale for MS (USE-MS): developing a patient based and patient reported outcome. *Mult Scler* 2012; 18: 1326–1333.
- Food and Drug Administration. Guidance for industry on patient-reported outcome measures: Use in medical product development to support labeling claims. Federal Register 2009; 74: 65132–65133.