J Rehabil Med 2011; 43: 944–950

ORIGINAL REPORT

RASCH ANALYSIS OF THE NOTTINGHAM EXTENDED ACTIVITIES OF DAILY LIVING SCALE

Roshan das Nair, PhD, Bryan J. Moreton, PhD and Nadina B. Lincoln, PhD

From the Institute of Work, Health and Organisations, University of Nottingham, Nottingham, UK

Background and purpose: **The Nottingham Extended Activities of Daily Living (EADL) Scale is frequently used in clinical practice and research in rehabilitation to assess patients' independence in activities of daily living. Summative scores are used for this purpose, but this is problematic because the EADL is an ordinal level measurement scale.**

Objectives: **To examine the fit of data to the Rasch model and to determine how the fit could be improved by making changes to the scale. The appropriateness of using total and subscale (Mobility, Kitchen, Domestic and Leisure) scores in determining change over time was evaluated.**

Methods: **EADL data (*n*=210 stroke patients, 55% male, age range 27–93 years) from a randomized trial of a Stroke family support organiser service were analysed using the Partial Credit model.**

Results: **Rasch analysis did not support the total scale as a unidimensional measure of activities of daily living. However, the subscales exhibited reasonable fit to the Rasch model following re-scoring and removal of items. Item 16 exhibited differential item functioning for age and item 22 differential item functioning for gender.**

Conclusion: **The results endorse the use and psychometric properties of the 4 EADL subscales, but not the total scale. Further work to corroborate these findings would be useful.**

Key words: Rasch analysis; stroke rehabilitation; disability; activities of daily living; outcome assessment.

J Rehabil Med 2011; 43: 944–950

Correspondence address: Roshan das Nair, B13 International House, Institute of Work, Health and Organisations, Jubilee Campus, University of Nottingham, Nottingham, NG8 1BB, UK. E-mail: roshan.nair@nottingham.ac.uk

Submitted January 5, 2011; accepted June 28, 2011

INTRODUCTION

One of the practical aims of stroke rehabilitation is to return patients to their pre-morbid environment with the ability to live as independently as possible. It is therefore important to assess their "instrumental" or extended activities of daily living (ADL). The Nottingham Extended Activities of Daily Living (EADL) (1) scale is an instrumental ADL scale, which is brief, easy to understand, self-administer, score, and interpret. The EADL has been validated as a postal questionnaire and is widely used in stroke services (2).

The EADL assesses the level of activity actually performed by a patient. Twenty-two activities are considered, which fall into 4 subscales: mobility, kitchen, domestic, and leisure activities. Responses are recorded using 1 of 4 options (*not at all*=0, *with help*=1, *on my own with difficultly*=2, *on my own*=3). The EADL was originally designed to function as a dichotomous scale [0011] with patients scored according to whether they were dependent on others [0] or independent [1] in their ADL (1). For this purpose, response options 0 and 1 [0] and 2 and 3 [1] were combined, respectively. However, in some studies (3), scoring has been based on the 4 response options [0123]. Currently, there is inconsistent use of both scoring methods in clinical practice. Summative subscale and total scores can be calculated from the EADL (4). However, there is evidence that suggests that the EADL is not a unidimensional construct, which questions the validity of a total summative score (4, 5). The dimensionality of the total EADL scale and subscales therefore needs to be evaluated further.

Studies with stroke patients have supported the validity (6) and test-retest reliability (1) of the EADL scale. When compared with other frequently employed scales of extended ADL, such as the Rivermead ADL (7), Hamrin Activity Index (8), and the Frenchay Activities Index (9), the EADL has better reported reliability and validity (10). The EADL has also been found to be superior to the Barthel Index (11) in validity and sensitivity in identifying disabilities in older adults (12). Nicholl et al. (5) tested the reliability and validity of the EADL with a group of 240 patients with multiple sclerosis. The subscales did not conform to Guttman scaling principles, but had satisfactory levels of internal consistency and test-retest reliability. Correlations between the subscales and quality of life measures were significant, but not particularly strong.

In clinical practice the EADL is used to assess progress or deterioration over time. This has been done by calculating change scores, but this is problematic given that the EADL is an ordinal level measurement tool, and is therefore bound by the limitations that this poses (13). Rasch analysis (14, 15) converts ordinal level measurements to interval level data, thereby facilitating the proper calculation of change scores and the use of parametric statistical methods. The Rasch model requires that a scale is unidimensional and functions in the same way across different groups of respondents (16). It therefore also functions as a stringent validity test for summative scores and cross-cultural performance.

 doi: 10.2340/16501977-0858
 ISSN 1650-1977

Tennant et al. (17) have used Rasch analysis to evaluate several frequently-used rehabilitation measures, such as the Functional Independence Measure (FIM) and the Mini-Mental State Examination (MMSE). The Rivermead Motor Assessment has also been analysed using the Rasch model (18).

The aims of this study were to explore the nature of the fit between data obtained from responses on the EADL scale (0–3 scoring) and the Rasch model and to explore ways in which the fit could be improved by making changes to the scale in terms of response categories and/or removal of items. The dimensionality of the total EADL scale and the subscales were also assessed.

METHODS

Patients admitted to hospitals in North Nottinghamshire with acute stroke (first or recurrent) were recruited for a trial to evaluate a stroke family support organiser service (SFO) (19). Patients were randomly assigned to an intervention group ($n = 104$) or a control group ($n = 106$) and assessed with the EADL 4 and 9 months after recruitment. Participants from both intervention and control groups were used in the current analysis. Questionnaires were sent by post, but if patients had difficulty completing questionnaires, an independent assessor who was blind to the group allocation visited them at their place of residence to assist with completion. When patients were unable to answer the questions, the EADL was obtained from discussions with their informal carer or, where appropriate, with allocated nurses in hospital or residential care homes. This data-set was chosen because the EADL was recorded on a 0–3 scale instead of using the 0011 scoring system.

Rasch analysis was conducted on data from 210 patients who completed the 4-month outcome assessment (19). This was deemed sufficient because a sample size of 150 patients provides 99% confidence that the item calibrations are within ±0.5 logits (20).

Rasch analysis

Rasch analysis was performed using the Rasch unidimensional measurement model (RUMM2020) (21) and was conducted in two stages. First, all items of the EADL were considered to evaluate the validity of the total scale (referred to as EADL-22) as a general measure of ADL. Secondly, the 4 individual subscales were analysed and reported separately (22). The Rasch model predicts responses to items as a logistic function of the level of trait expressed by an item and an individual (16). Dichotomous (14) and polytomous (15) versions of the model have been proposed and their application to analysis is described in more detail elsewhere (16, 23). To summarize, the primary focus of Rasch analysis is to assess the fit between data and the Rasch model to determine whether a scale meets its requirements, which is necessary for interval level data (24).

A likelihood ratio test was initially performed for the EADL-22 and each subscale to ascertain whether the Rating Scale (RS) model (15) or the Partial Credit (PC) model (25) was most suitable. Response thresholds were examined for each individual item to identify any disordered categories (16). Given misfit or marginal fit to the model, adjacent response categories were combined to resolve disordering (26), because this can improve the fit statistics (23). Invariance to the scale was assessed by examining the total χ^2 interaction term. A significant p-value at the 0.05 level is considered indicative of misfit between the data and model. Summary statistics for the item and person fit residuals were also calculated. Ideally the mean fit residual should be approximately 0 and the standard deviation should be approximately 1; deviation from these values signifies misfit. Fit residuals were also inspected at the individual level. Patients with fit residuals above or below 2.5 were deemed to be misfitting the model (23). χ^2 fit statistics were calculated for individual items to ensure that these met the model's expectations. A Bonferroni correction was applied to these tests to control for multiple testing.

With established scales, such as the EADL, it is desirable to avoid removing items as part of the analysis. However, it is sometimes necessary to do this to improve the fit to the Rasch model (27). Items and persons were only removed if: (*i*) there was evidence of individual item/person misfit; and (*ii*) removing the item/person improved the overall fit to the model.

The residual correlation matrices were examined for high positive correlations (0.3 and above), which indicates response dependency (28). Differential item functioning (DIF) was explored in two subgroups of patients formed on the basis of gender (male or female) and age (27–64, 65–74 and 75–93 years). Analysis of the variance with a Bonferroni correction was used to identify differences in responses. The assumption of unidimensionality was assessed using a specialized procedure in RUMM2020, based on work conducted by Smith (29). Principal components analysis (PCA) was used to establish positively and negatively loaded items in the residuals. These two subsets were then used to predict individual person values, which were compared with independent t-tests (16). If more than 5% of these t-tests are significant, then the scale is not unidimensional. A binomial confidence interval can be used if a scale fails the test (24). Finally, the person separation index (PSI) was calculated for each test and used as an indicator of the reliability of the fit statistics. A cut-off value of 0.7 was used as a minimum level of reliability (16).

RESULTS

Descriptive statistics

The sample was composed of 115 men (54.8%) and 95 women (45.2%). The age range was 27–93 years (mean = 69.17, standard deviation (SD) = 10.78 years) and the distribution of respondents across the 3 age groups was approximately equal (27–64 years, $n = 66$, 31.4%; 65–74 years, $n = 70$, 33.3%; 75–93 years, $n = 74$, 35.2%). Sixty-six (31.4%) patients completed the questionnaire unaided, 74 (35.2%) required assistance from an independent assessor, 66 (31.4%) questionnaires were completed by a carer, and for 4 (2.0%) patients the method of questionnaire completion was not recorded.

Rasch analysis of the Extended Activities of Daily Living-22

The likelihood ratio test was significant ($\chi^2 = 478.08$, degrees of freedom (df) = 41, $p < 0.001$) and so the PC model was used. All items except item 1 exhibited similarly disordered thresholds (see Fig. 1 for an example). For simplicity a global re-score was attempted (0112: *not at all, with help or difficulty, on my own*), but this left several items with disordered thresholds

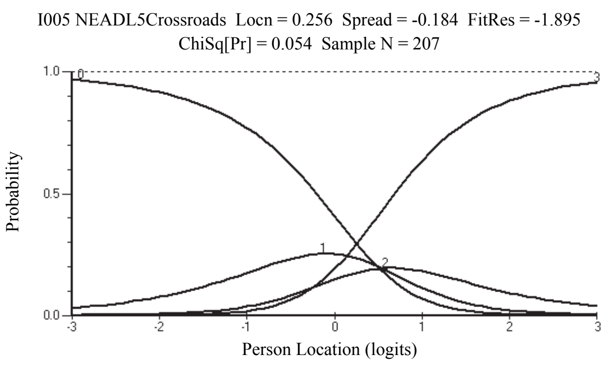

Fig. 1. Category probability curve for item 5.

(6, 8, 10, 11, 13, 16, 17, 19 and 22). Therefore, these items were dichotomized (0011: *not independent, independent*). Although re-scoring the disordered thresholds improved the fit to the Rasch model, there was still significant evidence of misfit. The mean fit residual for the items was –0.67 and the standard deviation (SD) was 1.46. The respective values for the person fit residuals (–0.34, 0.90) were more similar to the expected values of 0 and 1, which suggests less misfit to the model. The item-trait interaction statistic was highly significant ($\chi^2=123.58$, df=44, $p<0.001$), indicating a lack of invariance across the trait. At an individual level, 2 patients (Fit residuals: –2.58 and 4.11) and 3 items (item 8: Fit residual=–2.83, $\chi^2=18.16$, df=2, $p<0.001$; item 10: Fit residual=–2.68, $\chi^2=12.49$, df=2, $p<0.01$; item 11: Fit residual=–3.37, $\chi^2=15.92$, df=2, $p<0.001$) misfit the model. Examination of the residual correlation matrix revealed that there was evidence of response dependency: items 1–4 (0.33), 8–10 (0.38), 8–11 (0.35), 9–10 (0.38), 13 –14 (0.31), 13–16 (0.61), and 21–22 (0.49). Uniform DIF for gender was observed for items 5 and 22 and non-uniform DIF was observed for items 13 and 16. The EADL-22 scale also exhibited evidence of multidimensionality. Person estimates derived from the positively (8, 10, 11, 13, 14, 15 and 16) and negatively (1, 2, 3, 4 and 20) loaded items were compared using independent t-tests. Notice that the former subset is composed of only Domestic and Kitchen items, whereas the latter subset is made up primarily of Mobility items. Twenty-two of the 207 *t*-tests were significant at the 0.05 level (10.63%; confidence interval (CI): 7.70–13.60%). Considering these results, it was decided to discontinue this analysis and to focus on the subscales. Indeed, substantial changes to the EADL-22 would need to be made in order to achieve satisfactory fit to the model.

Rasch analysis of the Extended Activities of Daily Living subscales

The likelihood ratio tests for all subscales were statistically significant ($p<0.001$), which supported the use of the PC model. Three class intervals were used in the analyses, with approximately 50 patients in each, as recommended (30); the distribution of patients within the class intervals was monitored throughout.

The Mobility subscale (Items 1–6) exhibited disordered thresholds on 5 out of the 6 items. Items 2–5 were re-scored, 0112, and item 6 was re-scored, 0011. The overall fit statistics suggested some misfit to the model (Table I). Examination of the person fit residuals revealed that 4 responders were too deterministic in their answers (fit residuals=–2.90). Furthermore, items 1 and 3 were found to significantly misfit the model (item 1: fit residual=–3.16, $\chi^2=10.32$, df=2, $p<0.01$; item 3: fit residual=0.51, $\chi^2=11.75$, df=2, $p<0.01$). Item 1, in particular, had a high negative fit residual, which is indicative of redundancy. Removal of the 4 responders with high negative fit residuals and item 1 improved the overall fit to the model, as shown by the change in the χ^2 item-trait interaction term (see Table I). The revised 5-item subscale had no misfitting persons or items, no evidence of response dependency or DIF and was unidimensional. The PSI for the subscale was acceptable (0.80).

The Kitchen subscale (items 7–11) exhibited disordered thresholds for all but 1 (item 7) of the items. Items 8–10 were re-scored, 0112, and item 11 was re-scored, 0011, which resulted in ordered thresholds. Table I shows that the subscale initially misfit the Rasch model. Examination of the individual item fit statistics showed that items 7 and 8 deviated from expectations (item 7: fit residual=1.67, $\chi^2=10.72$, df=2, $p<0.01$; item 8: fit residual=–1.59, $\chi^2=11.36$, df=2, $p<0.01$). Item 7 had a positive fit residual, indicating marginal under-discrimination. Removal of item 7 improved fit to the model (see Table I). The revised subscale was free from response dependency and DIF and was unidimensional. The PSI was excellent and suitable for use at the individual level (0.91).

Items 12–16 form the Domestic subscale. All items in this subscale exhibited disordered thresholds, which were resolved by re-scoring items 12, 14, 15 and 16, 0112, and item 13, 0011. The subscale initially failed to fit the Rasch model. Items 12 and 13 produced significant χ^2 statistics (item 12: fit residuals=1.40, $\chi^2=11.03$, df=2, $p<0.01$; item 13: fit residual=–2.27, $\chi^2=10.09$, df=2, $p<0.01$). Item 12 had a positive fit residual indicating under discrimination and item 13 had a high negative fit residual, which is consistent with item redundancy. Removal of these items improved the fit to the model. There was no evidence of response dependency, but item 16 exhibited marginal non-uniform DIF for age. As Fig. 2 shows, the effect is primarily attributable to the 75–93 years group, which produced higher expected values at the first and third class intervals and a lower expected value at the second class interval. To explore the impact of DIF, separate person estimates were calculated before and after removal of item 16.

Table I. *Fit statistics for Extended Activities of Daily Living subscales*

Rasch analysis	Number of items χ	Item fit residual Mean (SD)	Person fit residual Mean (SD)	χ^2 (df)	*p*-value	Person Separation Index
Mobility (initial)	6	–0.45 (1.45)	–0.29 (0.86)	36.31 (12)	$p<0.001$	0.87
Mobility (removal of item 1)	5	–0.47 (0.61)	–0.30 (0.85)	17.54 (10)	$p=0.06$	0.80
Kitchen (initial)	5	–0.73 (1.37)	–0.34 (0.68)	35.71 (10)	$p<0.001$	0.87
Kitchen (removal of item 7)	4	0.19 (0.53)	–0.14 (0.74)	11.88 (8)	$p=0.16$	0.91
Domestic (initial)	5	–0.63 (1.35)	–0.36 (0.83)	37.92 (10)	$p<0.001$	0.86
Domestic (removal of items 12 & 13)	3	0.56 (0.73)	–0.70 (1.89)	5.76 (6)	$p=0.45$	0.82
Leisure	6	–0.41 (0.71)	–0.37 (0.59)	20.82 (12)	$p=0.053$	0.73

df: degree of freedom.

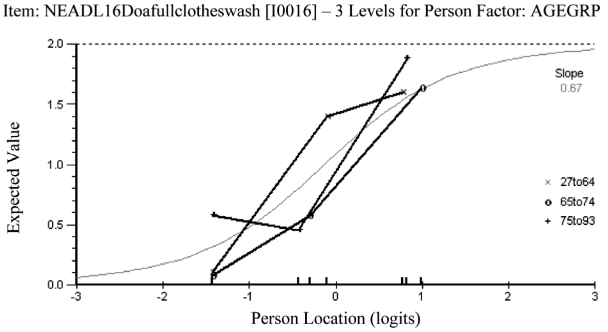

Fig. 2. Differential item functioning for the 3 age groups on item 16.

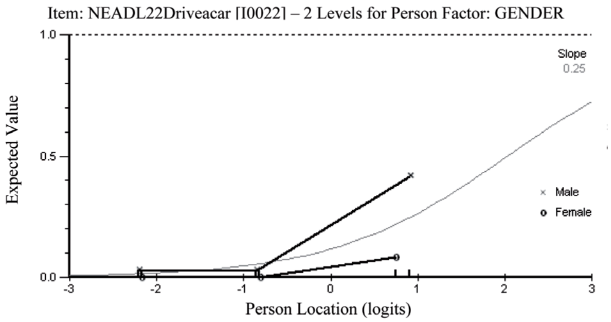

Fig. 3. Differential item functioning for males and females on item 22.

These estimates were then compared at an individual level and were found to differ by less than 0.3 logit on average, which indicates that removal of the item made little difference (22). Item 16 was therefore retained. Before items 12 and 13 were removed, the *t*-test procedure showed that the subscale was unidimensional. However, the test could not be performed following removal of these items, because the subsets were too small. There is however no evidence to refute that this is a unidimensional construct. A number of individuals presented with high negative fit residuals. However, removal of these patients did not improve the fit to the model. The PSI was acceptable for individual use (0.82).

All items on the leisure subscale (items 17–22) were disordered. Items 18, 20 and 21 were re-scored, 0112, and items 17, 19 and 22 were re-scored, 0011. The data exhibited reasonable fit to the model following the re-scoring procedure, with a low but acceptable PSI (0.73). All persons were within the acceptable fit residual range and none of the items significantly misfit the model. No evidence of response dependency was found in the person-item residual correlation matrix. However, item 22 exhibited marginal uniform DIF by gender. As shown in Fig. 3, the males produced higher expected values than the females, particularly at the third class interval. Person estimates derived before and after removal of item 22 were found to differ by less than 0.5 logits on average. This is a larger difference than observed previously with the Domestic subscale, but removal of item 22 was not found to improve fit to the Rasch model. Thus item 22 was retained. Independent *t*-tests for the positively and negatively loaded subsets showed that the subscale is unidimensional.

Kitchen and Domestic Subscale

The Kitchen and Domestic subscales contain a number of similar items, and so a final Rasch analysis was conducted on the combination of these two subscales. All items, except for item 14, exhibited disordered thresholds. Items 7, 8, 9, 12, 15 and 16 were re-scored, 0112, and items 10, 11 and 13 were re-scored, 0011. The data exhibited misfit to the Rasch model. One person had a high positive fit residual (3.15) and their removal improved the fit (Table II). Examination of the person-item residual correlation matrix showed that items 13 and 16 (0.36) exhibited local dependency. To resolve this, a super-ordinate *"washing"* item was created using the subtest procedure in RUMM2020. Only item 11 had a significant χ^2 statistic (fit residual = –3.33, χ^2 = 16.07, df = 2, $p < 0.001$). Removal of this item improved the fit, but the overall item-trait interaction was still significant, suggesting some misfit to the model. Following removal of item 11, all person fit statistics were within an acceptable range (± 2.5), there was no evidence of response dependency and no DIF. The fit residual for item 10 was high (–2.99), suggesting some misfit, but the χ^2 statistic was not significant. Removal of this item only marginally improved the fit.

Targeting

The person-item threshold distributions for the revised subscales showed that the mean location values were all negative (Fig. 4), which suggests that the patients were at a lower level of ADL than the items. Furthermore, quite large floor and ceiling effects were observed (Table III), particularly for Domestic activities.

DISCUSSION

The aim of this study was to explore the fit of data obtained from the Nottingham EADL scale with polytomous scoring (i.e. 0123) and the Rasch model. Previous studies (3) have

Table II. *Fit statistics for Kitchen/Domestic subscale*

Rasch Analysis	Number of items	Item fit residual Mean (SD)	Person fit residual Mean (SD)	χ^2 (df)	*p*-value	Person Separation Index
Kitchen/Domestic (initial)	10	–1.21 (1.56)	–0.50 (0.83)	63.83 (20)	$p < 0.001$	0.88
Kitchen/Domestic (removal of person)	10	–1.16 (1.50)	–0.48 (0.80)	54.19 (20)	$p < 0.001$	0.88
Kitchen/Domestic (subtest)	9	–1.11 (1.54)	–0.44 (0.79)	49.94 (18)	$p < 0.001$	0.88
Kitchen/Domestic (removal of item 11)	8	–1.03 (1.33)	–0.42 (0.75)	35.64 (16)	$p < 0.01$	0.86
Kitchen/Domestic (removal of items 11 and 10)	7	–0.90 (1.01)	–0.41 (0.70)	32.32 (14)	$p < 0.01$	0.85

df: degree of freedom.

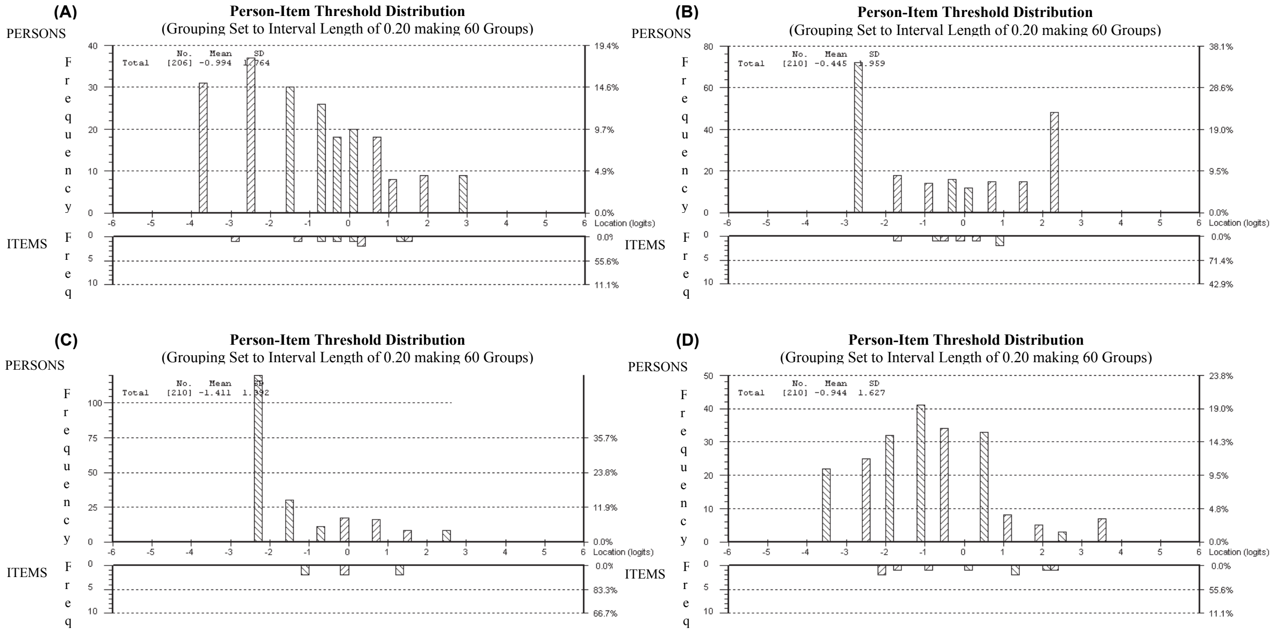

Fig. 4. Person-item threshold distribution for the Mobility (A), Kitchen (B), Domestic (C) and Leisure (D) subscales. SD: standard deviation.

calculated a total EADL score by summing the responses to all 22 items. Our analysis showed that: (*i*) most items presented with disordered thresholds; (*ii*) the EADL-22 was not unidimensional; and (*iii*) substantial alterations would be required to achieve fit to the Rasch model. Indeed, a study by Nicholl et al. (5) also found that the EADL-22 was not unidimensional with multiple sclerosis patients. These findings question the use of a total score, particularly when determining change over time, naturally or as a result of intervention. On the basis of these findings the 4 EADL subscales (mobility, kitchen, domestic and leisure activities) were analysed separately. The results confirmed that they were unidimensional, but, once again, the majority of items had disordered thresholds. Items were therefore either re-scored 0112 (*not at all, with help or difficulty, on my own)* or 0011 *(not independent, independent)* to resolve this issue. Items 1 *(Walk around outside)* and 13 *(Wash small items of clothing)* were found to have negative fit residuals suggesting item redundancy, and items 7 *(Manage to feed yourself)* and 12 *(Manage your own money when out)* had positive fit residuals indicating under-discrimination. Removal of these items improved fit to the Rasch model. In addition, 4 respondents were also removed during the analysis of the Mobility subscale, because they had high negative fit residuals. Following re-scoring and removal of these items and persons, all 4 subscales exhibited improved fit to the Rasch model with acceptable person separation indices (0.73–0.91). DIF was found for two items (16 and 22), but the effects were marginal and so they were retained. In a final analysis, the Kitchen and Domestic subscales were combined, but the fit between data and the model was weak.

Table III. *Percentage of floor and ceiling effects*

EADL subscale	Percentage floor effect	Percentage ceiling effect
Mobility	14.76	4.29
Kitchen	34.29	22.86
Domestic	57.14	3.81
Leisure	10.48	3.33

EADL: Extended Activities of Daily Living.

The EADL was originally designed to be scored 0011, where 0 is "not independent" and 1 is "independent" in ADL (1). It has since been scored 0123 *(not at all, with help, on my own with difficulty, on your own),* with the aim of improving its sensitivity to change. However, this scoring method consistently produced disordered thresholds. Response options 1 *(with help)* and 2 *(on my own with difficulty)* were particularly problematic. Both options indicate that a patient is able to perform a particular task with some problems. This is implied in response option 1 from the fact that they need help from another; whereas it is explicitly stated in response option 2. It is possible that patients were inconsistent in their interpretations of these categories and that some used "difficulty" as a proxy for physical help. For example, some may have considered supervision as "with help", while others may have reported it as "on my own with difficulty". Response options 1 and 2 were combined into a super-ordinate category *(with help or difficulty)* in an attempt to resolve this issue and retain a polytomous scale. However, a number of items still exhibited disordered thresholds and were therefore returned to the original dichotomous system. These changes to the scoring system should not have major implications for clinical practice or research purposes, because the EADL was always intended to be scored by collapsing response categories. If one wishes to use the Rasch-refined subscales (e.g. to convert the scores to an interval-level scale), then only the scoring technique would need to be altered.

The study had some limitations that need to be considered. To achieve satisfactory fit to the model a number of items had to be removed. However, we attempted to retain as many items as possible in the analyses, because it is undesirable to remove items from scales that are commonly used in clinical practice (23, 24). The Domestic subscale was affected the most, and was reduced to 3 items. The Rasch model is very stringent because it is designed to convert an ordinal scale to interval level measurement (24). This means that the scale may function adequately at the ordinal level without the necessity of item removal, so this would only be required when wishing to analyse results of the scale using parametric statistics (e.g. *t*-tests).

The targeting of the subscales was not particularly good, with quite a few patients producing floor effects. Low scores on a given subscale indicate that a patient is not able to perform respective ADL; whereas high scores signal that a patient is able to independently perform ADL. Currently there are insufficient items on the subscales to assess the level of trait in those with very low scores. However, sensitivity at this end of the ADL scale is perhaps not of paramount importance. From a clinical perspective, it is diagnostically useful to simply know that a patient is not able to perform a particular task. The targeting of the subscales should not have been a problem for the Rasch analysis. Indeed, even when the extreme respondents were not included in the analyses the sample sizes were sufficient to have approximately 95% confidence the item calibrations were within ± 0.5 logits (20).

A final limitation concerns the method of data collection. Data for this study was taken from a randomized controlled trial of a SFO intervention. Approximately half of the patients were allocated to receive the SFO and the remaining patients received only standard care. It is possible that this manipulation impacted their responses and affected the Rasch analysis. However, as reported in the original study (19), the scores produced on the EADL by these two groups were not significantly different. The primary effect of the SFO was improved knowledge concerning available support after stroke, stroke in general and ways to reduce the chance of another stroke. Therefore, it is unlikely that this manipulation had a large effect of the current analyses. Due to the nature of the patients' condition, it was sometimes necessary for a researcher to help the patient complete their questionnaire or to get the information from a carer. Once again, this might have had an impact on their data. Therefore, future research using a single mode of questionnaire administration (e.g. with help from a research assistant) would be useful to verify the current results.

In conclusion, Rasch analysis of the 4 subscales of the EADL generally supported their psychometric properties. However, re-scoring of the response categories and removal of 4 items was undertaken to improve the fit between data and the Rasch model. DIF was observed on two items, but these were not large effects. The use of the total (EADL-22) scale as a general measure of ADL was not supported, because the scale was not unidimensional.

REFERENCES

1. Nouri FM, Lincoln NB. An extended activities of daily living scale for stroke patients. Clin Rehabil 1987; 1: 301–305.
2. Gladman JRF, Lincoln NB, Adams SA. Use of the extended ADL scale with stroke patients. Age Ageing 1993; 22: 419–424.
3. Parker CJ, Gladman JRF, Drummond AER, Dewey ME, Lincoln NB, Barer D, et al. A multi-centre randomised controlled trial of leisure therapy and conventional occupational therapy after stroke. Clin Rehabil 2001; 15: 42–52.
4. Harwood RH, Ebrahim S. The validity, reliability and responsiveness of the Nottingham Extended Activities of Daily Living Scale in patients undergoing total hip replacement. Disabil Rehabil 2002; 24: 371–377.
5. Nicholl CR, Lincoln NB, Playford ED. The reliability and validity of the Nottingham Extended Activities of Daily Living Scale in patients with multiple sclerosis. Multiple Sclerosis 2002; 8: 372–376.
6. Lincoln NB, Gladman JRF. The extended ADL scale: a further validation. Disabil Rehabil 1992; 14: 41–43.
7. Whiting S, Lincoln NB. An ADL assessment for stroke patients. Br J Occupat Ther 1980; 43: 44–46.
8. Hamrin E. One year after stroke: a follow-up of an experimental study. Scand J Rehabil Med 1982; 14: 111–116.
9. Holbrook M, Skilbeck CE. An activities index for use with stroke patients. Age Ageing 1983; 12: 166–170.
10. Chong DK. Measurement of instrumental activities of daily living in stroke. Stroke 1995; 26: 1119–1122.
11. Mahoney FI, Barthel B. Functional evaluation: the Barthel Index. Maryland State Med J 1965; 14: 56–61.
12. Yohannes AM, Roomi J, Waters K, Connolly MJ. A comparison of the Barthel Index and Nottingham Extended Activities of Daily Living Scale in the assessment of disability in chronic airflow limitation in old age. Age Ageing 1998; 27: 369–374.
13. Siegel S. Nonparametric statistics. The American Statistician 1957; 11: 13–19.
14. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago; 1960.
15. Andrich D. Rating formulation for ordered response categories. Psychometrika 1978; 43: 561–573.
16. 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? Arthrit Rheumatol 2007; 57: 1358–1362.
17. Tennant A, Penta M, Tesio L, Grimby G, Thonnard J-L, Slade A, et al. Assessing and adjusting for cross cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the rasch model: the pro-esor project. Med Care 2004; 42 Suppl 1: S1–S37.
18. van de Winckel A, Feys H, Lincoln NB, De Weerdt W. Assessment of arm function in stroke pateints: rivermead motor assessment arm section revised with Rasch analysis. Clin Rehabil 2007; 21: 471–479.
19. Lincoln NB, Francis VM, Lilley SA, Sharma JC, Summerfield M. Evaluation of a stroke Family Support Organiser: a randomized controlled trial. Stroke 2003; 34: 116–121.
20. Linacre JM. Sample size and item calibration stability. Rasch Meas Transact 1994; 7: 328.
21. Andrich D, Lyne A, Sheridan B, Luo G. RUMM 2020. Perth: RUMM Laboratory; 2003.
22. Pallant JF, Keenan A-M, Misajon RA, Conaghan PG, Tennant A. Measuring the impact and distress of Osteoarthritis from the patients' perspective. Health Qual Life Outcomes 2009; 7: 37.
23. Pallant JF, 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.

24. Shea TL, Tennant A, Pallant JF. Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). BMC Psychiatry 2009; 9.
25. Masters GN, Wright BD. The Partial Credit Model. In: van der Linden WJ, Hambleton RK, editors. Handbook of modern item response theory. New York: Springer; 1997.
26. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: An introduction and illustrative example. Int J Nurs Studies 2009; 46: 380–393.
27. Mills RJ, Young CA, Nicholas RS, Pallant JF, Tennant A. Rasch analysis of the Fatigue Severity Scale in Multiple Sclerosis. Multiple Sclerosis 2009; 15: 81–87.
28. Ramp M, Khan F, Misajon RA, Pallant JF. Rasch analysis of the Multiple Sclerosis Impact Scale (MSIS-29). Health Qual Life Outcomes 2009; 7: 58.
29. Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal components analysis of residuals. J Applied Meas 2002; 3: 205–231.
30. Lundgren-Nilsson Å, Tennant A, Grimby G, Sunnerhagen KS. Cross-diagnostic validity in a generic instrument: an example from the Functional Independence Measure in Scandinavia. Health Qual Life Outcomes 2006; 4: 55–62.

APPENDIX I. Raw score, Rasch score and standard error of the revised Extended Activities of Daily Living subscales

Subscale	Raw score	Rasch score	Standard error
Mobility	0	–3.67	1.57
	1	–2.42	1.13
	2	–1.44	0.89
	3	–0.80	0.78
	4	–0.29	0.73
	5	0.17	0.72
	6	0.64	0.75
	7	1.20	0.82
	8	1.94	1.00
	9	2.87	1.38
Kitchen	0	–2.61	1.38
	1	–1.69	1.01
	2	–0.94	0.85
	3	–0.36	0.78
	4	0.17	0.78
	5	0.72	0.83
	6	1.41	0.98
	7	2.29	1.34
Domestic	0	–2.39	1.41
	1	–1.43	1.03
	2	–0.66	0.88
	3	–0.04	0.85
	4	0.60	0.90
	5	1.46	1.07
	6	2.48	1.45
Leisure	0	–3.42	1.34
	1	–2.59	1.00
	2	–1.88	0.89
	3	–1.20	0.88
	4	–0.40	0.89
	5	0.43	0.88
	6	1.17	0.86
	7	1.84	0.89
	8	2.58	1.01
	9	3.46	1.36