Use of the functional assessment measure (FIM+FAM) in head injury rehabilitation: a psychometric analysis

Carol A Hawley, Robert Taylor, Deborah J Hellawell, Brian Pentland

Centre for Health Services Studies, University of Warwick, Coventry, CV4 7AL, UK
C A Hawley

Department of Clinical Neurosciences, Western General Hospital, Edinburgh EH4 2XU, UK
R Taylor

Rehabilitation Studies Unit, Astley Ainslie Hospital, Edinburgh EH9 2HL, UK
D J Hellawell
B Pentland

Correspondence to:
Carol Hawley, Senior Research Fellow, Centre for Health Services Studies, University of Warwick, Coventry, CV4 7AL. email
c.a.hawley@warwick.ac.uk

Abstract

Objectives—The drive to measure outcome during rehabilitation after brain injury has led to the increased use of the functional assessment measure (FIM+FAM), a 30 item, seven level ordinal scale. The objectives of the study were to determine the psychometric structure, internal consistency, and other characteristics of the measure.

Methods—Psychometric analyses including both traditional principal components analysis and Rasch analysis were carried out on FIM+FAM data from 2268 assessments in 965 patients from 11 brain injury rehabilitation programmes.

Results—Two emergent principal components were characterised as representing physical and cognitive functioning respectively. Subscales based on these components were shown to have high internal consistency and reliability. These subscales and the full scale conformed only partially to a Rasch model. Use of raw item ratings, as opposed to transformed ratings, to produce summary scores for the two subscales and the full scale did not introduce serious distortion.

Conclusion—The full FIM+FAM scale and two derived subscales have high internal reliability and the use of untransformed ratings should be adequate for most clinical and research purposes in comparable samples of patients with head injury.

(*J Neurol Neurosurg Psychiatry* 1999;**67**:749–754)

Keywords: functional assessment measure; brain injury; psychometrics; principal components; Rasch

Acute brain injury commonly results in a combination of physical, cognitive, and behavioural consequences which require labour intensive, and often protracted, rehabilitation.[1][2] The development of specialist rehabilitation services for patients with brain injury has been accompanied by pressure for clinical audit, not least for economic reasons.[3] In some countries, particularly the United States where centres offering rehabilitation proliferated in the 1980s, there have been several attempts to standardise instruments designed to evaluate rehabilitation programmes.[4–6] The functional independence measure (FIM) has gained increasing popularity as an outcome measure for general use in medical rehabilitation, including rehabilitation after head injury.[7][8] The FIM scores 18 functional activities on a seven level scale. In view of the prominence of communicative, cognitive, and behavioural disturbances after brain injury a further 12 items considering those issues were added to the FIM to construct the functional assessment measure.[6] It has become accepted custom to use the abbreviation FIM+FAM for the complete 30 item functional assessment measure.[6] Although the FIM+FAM has been increasingly adopted to measure outcome in rehabilitation after brain injury, its psychometric properties have not been investigated extensively.[6][9][10]

Opinion has varied over the years concerning the extent to which various psychological and behavioural ratings approximate to points on true interval scales of measurement, and hence to what extent it is appropriate to subject them to various mathematical operations and parametric statistical procedures.[11–14] The availability of user friendly computerised versions of elegant Rasch statistical models[15] has no doubt contributed to the current revival of this debate.

We present an investigation of the psychometric properties of FIM+FAM ratings from a large multicentre population of patients undergoing rehabilitation after traumatic brain injury. This includes analysis of the principal component structure of the scale; evaluation of derived subscales; and determination of the extent to which the whole scale and the subscales conform to a Rasch model of an ideal homogeneous measurement instrument. In addition, consideration is given to the effects of using various indices derived from raw ratings compared with using the raw ratings in computing summary scores, or in examining profiles of rated functioning on individual scale items.

Method

PATIENTS

The Department of Health awarded grants to 10 sites in England to enhance their existing brain injury rehabilitation services with the requirement that they contribute data to the National Traumatic Brain Injury (NTBI) study.[16] The Centre for Health Services Studies, University of Warwick, was given the task of coordinating and evaluating the data. The 10 collaborating centres are listed in the acknowledgements. Patients were registered over a period of 3 years, from 1992 to 1995. One of the principal measures used in the NTBI study was the FIM+FAM which was scored at about 3 months, 18 months, and, wherever possible,

Table 1 Patient details

Location	No of patients	No of males	Age			No of assessments
			Mean	(SD)	Range	
AAH*	313	239	38.1	(17.2)	14 to 78	934
Warwick cohort	652	494	34.4	(13.0)	16 to 72	1334
Total	965	733	35.6	(14.6)	14 to 78	2268

AAH=Astley Ainslie Hospital, Edinburgh.

Table 2 Rotated principal component loadings, with higher loading for each item italicised

Item	Component 1	Component 2
Bed transfer	*0.888*	0.383
Toilet transfer	*0.887*	0.390
Car transfer	*0.875*	0.405
Toileting	*0.851*	0.453
Bath transfer	*0.846*	0.431
Walking/wheelchair	*0.835*	0.407
Dressing lower	*0.823*	0.483
Bowel	*0.807*	0.440
Dressing upper	*0.802*	0.495
Feeding	*0.800*	0.441
Stairs	*0.795*	0.401
Bladder	*0.784*	0.483
Bathing	*0.769*	0.531
Grooming	*0.752*	0.567
Swallowing	*0.738*	0.377
Speech intelligibility	*0.630*	0.571
Problem solving	0.352	*0.863*
Adjustment to limitations	0.363	*0.841*
Memory	0.353	*0.837*
Attention	0.418	*0.818*
Employability	0.371	*0.787*
Social interaction	0.451	*0.786*
Safety judgement	0.514	*0.782*
Orientation	0.482	*0.780*
Comprehension	0.460	*0.769*
Expression	0.495	*0.766*
Emotion	0.368	*0.736*
Reading	0.466	*0.734*
Writing	0.507	*0.734*
Community mobility	0.563	*0.675*

at 3 years postregistration. The numbers of patients from each site who were assessed ranged from 23 to 124, and data from 652 patients were included in this analysis. The nature of the rehabilitation programme varied between centres but only information from cases of traumatic brain injury is included. These patients constitute the Warwick cohort.

The Scottish Brain Injury Rehabilitation Service, Edinburgh (SBIRSE) provides early inpatient rehabilitation to patients with brain injury from throughout Scotland (population about 4.5 million). There are 20 beds to which patients are admitted from acute surgical and medical units after traumatic and non-traumatic brain injury. The inpatient rehabilitation is multiprofessional with weekly case conferences on all patients to plan, review, and adjust individual programmes and to plan discharge arrangements. The FIM+FAM is scored within 48 hours of admission and in the week before discharge from the unit, and at monthly intervals in those with a prolonged duration of stay. Again, only the results from patients with traumatic brain injury are included in this study. The patient details of both the Warwick and Edinburgh cohorts are summarised in table 1.

STATISTICAL ANALYSIS

Principal components analysis (PCA) is a well established method of investigating the structure of rating scales[17]; PCA with varimax rotation was carried out on data from all assessments. All non-missing data were included by using pairwise as opposed to casewise deletion of missing data. Cronbach's α was computed in the standard manner as a measure of the internal consistency and reliability of the whole scale and the subscales derived from PCA. Pearson's correlation coefficients were computed in the standard manner.

Rasch models and analyses allow the properties of ordinal scales to be studied after their conversion to ordered, unidimensional interval scales. The conceptual and mathematical bases of Rasch analysis have been the subject of several reviews.[13 15] Rasch procedures, in the context of functional activity scales such as the FIM+FAM, should provide clarification of the structure of scales and the characteristics of their constituent items. Raw item ratings are transformed into scores which can be considered points on true interval scales of measurement where individual scale items are adjusted for differing "difficulty levels" and are expressed in a common metric (logits or log odds units) across items. Estimates are derived of difficulty levels of scale item ratings and ability levels of people that are relatively independent

of each other, and hence relatively independent of the particular patient sample studied.

Rasch analyses were carried out with the Winsteps/Bigsteps Rasch model program[15] using the partial credit version (specified using the instruction "groups=0") of the rating scale model in the program, whereby the rating levels 1 to 7 are not considered equivalent across items (each item being analysed as a small rating scale in its own right within a larger one comprising all relevant items). All non-missing data were included according to standard (default) program procedures.

Various transformations of raw ratings for use in computing summary scores were derived via traditional psychometric procedures and Rasch analysis. All analyses other than Rasch were conducted using SPSS Version 7.5.1.[18]

Results

PRINCIPAL COMPONENTS ANALYSIS

Two principal components with eigenvalues >1 emerged. Table 2 gives details of rotated component loadings on these two components, with higher loading (each of which is at least 0.63) for each item italicised. It can be seen that the 16 items loading most highly on the first component essentially reflect physical functioning, and the 14 items loading most highly on the second reflect aspects of cognitive, language, and psychosocial functioning. This factorial structure seems encouragingly coherent and comprehensible, and the components can be appropriately characterised as physical and cognitive respectively. The percentages of variance accounted for by the two factors were 77.1 and 6.5 respectively before rotation (and 44.0 and 39.6 respectively after), 83.6 in total.

Items were grouped into two subscales on the basis of the principal components analysis. Cronbach's α based on raw ratings was 0.99 for

the 16 item physical subscale, 0.98 for the 14 item cognitive subscale, and 0.99 for the whole 30 item scale. These are highly acceptable levels. (The fact that α is partially dependent on number of constituent items in a scale explains the marginally higher figure for the whole scale than for one of the subscales.) Further item analysis indicated that no increase in α was obtained by omitting any given item within each of the subscales or the scale as a whole.

RASCH ANALYSIS

Two statistics produced within the Rasch analyses, concerning the extent to which the data conform to or fit a Rasch model of an ideal homogeneous measurement instrument, are particularly relevant here: "Infit" is an information weighted fit statistic in which unusually high values indicate noise in the data; "Outfit" is an outlier sensitive statistic in which unusually high values indicate unexpected outlying ratings—that is, ratings of subjects on any given item that are much higher or lower than would be expected on the basis of the difficulty levels of item ratings and the estimated ability levels of subjects. Unusually low values of either statistic suggest dependency or redundancy in the data. Desirable ranges of these statistics are here taken as 0.7 to 1.3 as suggested by Linacre and Wright.[15] Here we focus on how well items (as opposed to persons) fit the model.

For the 16 item physical subscale, three items (speech intelligibility, stairs, swallowing) had infit values above the desirable range, and four (bladder, speech intelligibility, stairs, swallowing) had outfit values above the desirable range. For the 14 item cognitive subscale, two items (community mobility, emotion) had infit values above the desirable range, and three (community mobility, emotion, reading) had outfit values above the desirable range. When the scale as a whole was analysed as a single scale, five of the 30 items (emotion, reading, speech intelligibility, stairs, swallowing) had infit values above the desirable range, and four (emotion, reading, speech intelligibility, stairs) had outfit values above the desirable range. In each of the three analyses, some items had values below the desirable range for infit, or outfit, or both but these items are not listed here as the implication of a degree of redundancy in scale items should not be considered a serious problem, and might be considered an advantage, with a clinical instrument such as this. In these circumstances, it is not surprising that fit statistics for persons (as opposed to items) indicated that departures from the model were common for individual persons in the sample.

Indices of person separation and item separation[15]—that is, the extent to which a scale is able to discriminate various different levels of ability or performance, and associated reliabilities—were acceptably high for each subscale and for the scale as a whole. Despite the imperfect fit of these data to the model, it remains the case that the properties of raw score transformations derived from Rasch analysis (including numerical measurement properties, and inherent adjustment for varying item difficulty) may confer important advantages for purposes of analysis or interpretation of data; this possibility is considered further below.

RAW RATINGS VERSUS ALTERNATIVE DERIVED INDICES OF FUNCTION

If, as previously stated, the raw item ratings do not constitute true interval scale scores, they are not suitable for simple arithmetic procedures such as addition. Also, the combination of raw ratings on items of differing weights or difficulty levels to produce a summary score may be misleading and an inaccurate reflection of the overall functional status. Hence it is important to consider how the use of untransformed raw ratings compares with the use of appropriate alternative derived indices of function.

Thus various indices of function were calculated for each of the two subscales and for the whole 30 item scale for each patient assessment. These were (*a*) the mean of the raw ratings of the patient on the relevant items; (*b*) the median of the raw ratings of the patient on those items; (*c*) the mean of the ratings after transformation of raw ratings on each item to standardised scores, with mean of 0 and SD of 1, based on the means and SDs of all non-missing raw ratings for that item in the whole sample; (*d*) the mean of the ratings after transformation of raw ratings on each item to normalised scores in a forced normal distribution, with mean of 0 and SD of 1, based on the percentile distribution of all non-missing raw ratings for that item in the whole sample; (*e*) the mean of the ratings after transformation of raw ratings on each item to Rasch scaled scores derived empirically from the distributions of ratings in this patient sample (given by the average measure for each rating level on each item in Winsteps/Bigsteps), henceforth referred to as RaschAM; and (*f*) the mean of the ratings after transformation of raw ratings on each item to more theoretically derived Rasch scaled scores that are considered largely independent of the particular sample from which they are derived (given by the "score-to-measure at" category in Winsteps/Bigsteps), henceforth referred to as Rasch StMaC. Finally (*g*), a factor score was computed in the standard manner via principal components analysis for each of the two subscales: such scores take into account principal component loadings for every item on each factor, rather than simply assigning each item to one subscale according to which factor it loads most highly upon. Most of these transformations take some account (to varying degrees) of inherent differences in item difficulties: the Rasch StMaC transformation might be expected to be the least sample dependent and most generalisable transformation.

All these summary scores were calculated only for patient assessments with no missing data for the scale in question: the numbers of assessments were 1816 for the physical subscale, 1709 for the cognitive subscale, and 1572 for the whole scale. In practice, given the high levels of internal consistency found, it would be

Table 3 Correlations (Pearson's r) between mean raw rating and other numerical indices of functioning for each of the subscales identified by principal components analysis, and for the scale as a whole

	Mean raw physical	Mean raw cognitive	Mean raw whole
Median raw	0.989	0.987	0.980
Mean standardised	1.000	1.000	1.000
Mean forced-normal	0.968	0.983	0.981
Mean Rasch-scaled AM	0.991	0.993	0.994
Mean Rasch-scaled StMaC	0.993	0.991	0.989
Factor score	0.859	0.881	—

Figure 1 Scatterplot of mean Rasch versus raw score on the physical subscale.

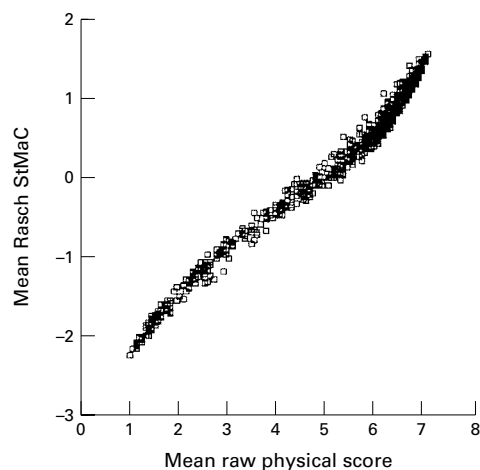*Figure 2 Scatterplot of mean Rasch versus raw score on the cognitive subscale.*

Figure 3 Scatterplot of mean Rasch versus raw score on the 30 item scale.

reasonable to prorate mean ratings for the subscales or whole scale where only one or two item ratings are missing (simply by computing the mean of available ratings).

Table 3 shows correlations (Pearson's *r*) between mean raw rating and these other numerical indices of functioning for each of the two subscales and for the scale as a whole. The correlations between mean raw ratings and the other indices of principal interest (those in the first five rows of the table) are very high, and indicate a very high proportion of shared variance (equal to *r²*). The perfect correlations between mean raw and standardised scores are unsurprising given that the standardised score for any item is a linear transformation of the raw rating. Figures 1–3 show scatterplots of mean Rasch StMaC score against mean raw rating for each of the subscales and the scale as a whole. These give a visual impression of the close correspondence between the mean raw ratings and scores based on the least sample dependent and most generalisable transformation. The slight curvilinearity of the relations seen here is not uncommon when appropriate psychometric transformations are applied to raw scores or ratings of various kinds.

Rasch scaled (StMaC) logit equivalents for each raw rating on each item, based on analysis of the whole 30 item scale, are presented in table 4. These figures can be used in computation of summary scores or in plotting profiles of performances of patients or groups of patients on individual items. Use of the whole 30 item scale in deriving these figures will tend to provide the most consistent adjustment for item difficulty 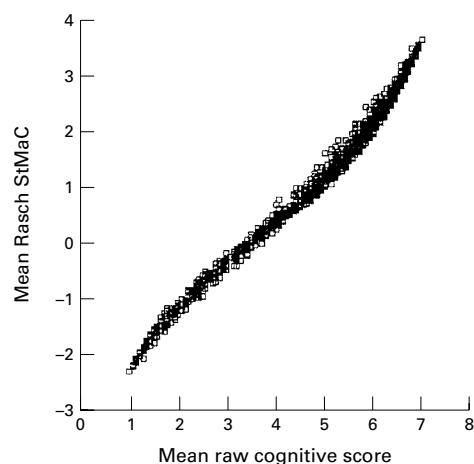across items which appear in different subscales: such a procedure is theoretically questionable given that subscales have been identified, but it seems from other analyses that this issue is unlikely to be of practical significance. Details of Rasch scaled transformations for the separate subscales are available from Robert Taylor. For information, table 5 presents details of raw ratings for each item, and of mean raw ratings for the two subscales and the scale as a whole.

For each of the subscales, and for the scale as a whole, Cronbach's α was calculated using each relevant transformation of raw ratings (standardised, forced normal, and the two Rasch transformations): the resulting figures were no higher than those based on analyses of raw ratings.

In some research or clinical contexts it may be appropriate to consider differences between levels of functioning on the two subscales. The difference (calculated for 1572 assessments) based on mean raw ratings correlated 0.933 with the difference based on median measures; 0.999 with the difference based on standardised scores; 0.908 with the difference based on forced normal transformations; 0.959 with the difference based on Rasch AM transforma-

Table 4   Rasch-scaled (StMaC) logit equivalents for each raw rating on each item

Item	Raw rating						
	1	2	3	4	5	6	7
Bed transfer	−2.31	−1.27	−0.80	−0.48	−0.16	0.31	1.37
Toilet transfer	−2.20	−1.24	−0.77	−0.45	−0.13	0.33	1.37
Car transfer	−1.87	−0.96	−0.52	−0.22	0.09	0.54	1.59
Toileting	−1.82	−0.90	−0.51	−0.26	0.00	0.36	1.21
Bath transfer	−1.97	−1.08	−0.60	−0.23	0.19	0.83	2.10
Walking/wheelchair	−1.70	−0.98	−0.65	−0.40	−0.09	0.55	2.25
Dressing lower	−2.41	−1.17	−0.59	−0.21	0.17	0.69	1.70
Bowel	−1.84	−1.10	−0.76	−0.53	−0.30	0.03	0.75
Dressing upper	−2.95	−1.61	−0.98	−0.56	−0.14	0.44	1.49
Feeding	−2.52	−1.68	−1.25	−0.92	−0.53	0.11	1.29
Stairs	−0.75	−0.28	−0.02	0.19	0.44	0.98	2.71
Bladder	−1.78	−0.94	−0.58	−0.33	−0.10	0.23	0.90
Bathing	−2.25	−1.20	−0.66	−0.22	0.28	0.98	2.13
Grooming	−2.52	−1.41	−0.88	−0.48	−0.04	0.61	1.73
Swallowing	−3.69	−2.49	−1.88	−1.45	−1.03	−0.49	0.50
Speech intelligibility	−3.24	−2.00	−1.19	−0.60	−0.06	0.67	2.13
Problem solving	−1.71	−0.44	0.27	0.82	1.44	2.38	3.96
Adjustment to limitations	−2.19	−0.71	−0.02	0.48	1.11	2.34	4.25
Memory	−2.36	−0.85	0.02	0.67	1.37	2.49	4.38
Attention	−3.17	−1.45	−0.42	0.33	1.11	2.22	3.93
Employability	−1.37	−0.14	0.61	1.32	2.24	3.66	5.76
Social interaction	−3.07	−1.37	−0.49	0.06	0.62	1.56	3.35
Safety judgement	−2.34	−0.83	−0.10	0.40	0.95	1.85	3.44
Orientation	−2.04	−0.92	−0.36	0.04	0.46	1.08	2.30
Comprehension	−3.18	−1.70	−0.94	−0.41	0.15	0.99	2.46
Expression	−2.67	−1.24	−0.56	−0.08	0.43	1.25	2.77
Emotion	−2.71	−1.13	−0.21	0.48	1.24	2.59	4.84
Reading	−2.19	−1.26	−0.76	−0.31	0.37	1.61	3.38
Writing	−1.86	−0.91	−0.32	0.20	0.83	1.74	3.19
Community mobility	−1.23	−0.20	0.31	0.69	1.12	1.80	3.13

Table 5   Means and SDs of raw ratings for each item, and of mean raw ratings for the two subscales and the scale as a whole; and numbers of assessments for each item and scale

Item	Mean	SD	No of assessments
Bed transfer	5.78	2.03	2250
Toilet transfer	5.77	2.04	2250
Car transfer	5.75	2.08	1953
Toileting	5.68	2.19	2251
Bath transfer	5.48	2.09	2242
Walking/wheelchair	5.60	2.07	2244
Dressing lower	5.56	2.10	2255
Bowel	5.91	2.09	2192
Dressing upper	5.79	1.94	2256
Feeding	5.98	1.84	2255
Stairs	5.15	2.37	2184
Bladder	5.78	2.18	2194
Bathing	5.44	2.06	2250
Grooming	5.69	1.98	2252
Swallowing	6.30	1.60	2228
Speech intelligibility	5.70	1.84	2228
Problem solving	4.42	2.19	2209
Adjustment to limitations	4.68	2.08	2178
Memory	4.55	2.04	2221
Attention	4.84	1.93	2208
Employability	4.03	2.06	1947
Social interaction	5.13	1.99	2206
Safety judgement	4.85	2.13	2219
Orientation	5.28	2.17	2210
Comprehension	5.52	1.90	2242
Expression	5.28	2.03	2242
Emotion	4.67	1.94	2181
Reading	5.26	1.92	2190
Writing	4.99	2.11	2168
Community mobility	4.84	2.32	1864
Physical subscale	5.88	1.82	1816
Cognitive subscale	5.07	1.84	1709
Whole scale	5.57	1.73	1572

tions; 0.818 with the difference based on Rasch StMaC transformations; and 0.994 with the difference based on factor scores. These correlations are again reassuringly high.

Discussion

The results of this study of FIM+FAM assessments from a large population of patients with traumatic brain injury indicate that the FIM+FAM has a highly acceptable level of internal consistency and reliability. This applies both to the full 30 item scale and the 16 item motor and 14 item cognitive subscales derived by PCA. Hall *et al*[6] reported similar findings but without presenting the relevant data.

The FIM+FAM data from this sample did not conform particularly well to a Rasch model, although the departures from the model were not extreme. The imperfect fit of the FIM+FAM scale and subscales to a Rasch model is, however, not surprising in the present context, and does not indicate that the scale and subscales are fundamentally flawed. Closeness of fit to a Rasch model depends to some extent on the degree of homogeneity of the sample studied, as well as the extent to which the items of the scale are essentially unidimensional. The PCA clearly indicates that the full 30 item scale in this sample is not unidimensional. Similarly, the extent to which Rasch scaled (StMaC) raw rating transformations are independent of the particular sample of patients from which they were derived, and the extent to which they can be generalised, will depend on the extent to which study samples are homogeneous and representative of other samples or populations. Clearly it would be unreasonable to expect the relative difficulties of, for example, stairs and swallowing to be very similar across patients with very different patterns of neurological disability in a brain injury rehabilitation unit. Patients with traumatic brain injury are notoriously heterogeneous in the range and extent of their disabilities. Dickson and Kohler[19] summarise other limitations in applying Rasch analytical procedures to data obtained using this type of scale, especially across different patient samples. We do not think that it would be appropriate to attempt or recommend removal or modification of scale items on the basis of this analysis in one (albeit large) sample of patients after brain injury.

In practice, Rasch transformed scores may be difficult to interpret clinically without reference to rated functioning in other appropriate samples of patients, thereby approximating a more traditional psychometric approach. Also, graphic profiles of performance on individual items based on the Rasch transformations presented in table 4 may not be superior in all respects to profiles produced using raw ratings: the inherent adjustment for differences in item difficulty may aid interpretation of the pattern of patients' relative strengths and limitations around the middle of the item score ranges, but hamper interpretation nearer the extremes in that, for example, the representation of maximal performance (all raw ratings equal to 7) or minimal performance (all ratings equal to 1) will consist of a markedly jagged or uneven profile as a result of the differences in item difficulties. The raw ratings are more soundly behaviourally (as opposed to theoretically) anchored.

In any branch of applied science, the worth of measures or models tends to depend upon their utility rather than mathematical purity or elegance (even though the two may coincide);

approximations tend to be the norm rather than the exception. Our results suggest that treating FAM raw ratings as good and useful approximations to points on interval scales of measurement, and treating them arithmetically to characterise levels of functioning on subscales or the scale as a whole, is justifiable and will not introduce serious distortion; so that the inconvenience of transforming all raw ratings before carrying out any analysis or interpretation of data can reasonably be avoided.

1 Scottish Home and Health Department. *Medical rehabilitation: the pattern for the future.* Edinburgh: HMSO, 1972.
2 British Society of Rehabilitation Medicine. *Rehabilitation after traumatic brain injury.* London: British Society of Rehabilitation Medicine, 1998.
3 McGregor K, Pentland B. Head injury rehabilitation in the UK: an economic perspective. *Soc Sci Med* 1997;**45**:295–303.
4 Harvey RF, Jellinek HM. Functional performance assessment: a program approach. *Arch Phys Med Rehabil* 1981;**62**:456–60.
5 Rappaport M, Hall KM, Hopkins K, *et al*. Disability rating scale for severe head trauma: coma to community. *Arch Phys Med Rehabil* 1982;**63**:118–23.
6 Hall KM, Hamilton BB, Gordon WA, *et al*. Characteristics and comparisons of functional assessment indices: disability rating scale, functional independence measure, and functional assessment measure. *J Head Trauma Rehabil* 1993;**8**:60–74.
7 Granger CV, Hamilton BB, Keith RA. Advances in functional assessment for medical rehabilitation. *Topics in Geriatric Medicine* 1986;**1**:59–74.
8 Granger CV, Divan N, Fiedler RC. Functional assessment scales: a study of persons after traumatic brain injury. *Am J Phys Med Rehabil* 1995;**74**:107–13.
9 Donaghy S, Wass PJ. Interrater reliability of the functional assessment measure in a brain injury rehabilitation program. *Arch Phys Med Rehabil* 1998;**79**:1231–6.
10 Tesio L, Cantagallo A. The functional assessment measure (FAM) in closed traumatic brain injury outpatients: a Rasch-based psychometric study. *Journal of Outcome Measurement* 1998;**2**:79–96.
11 Stevens SS. Mathematics, measurement, and psychophysics. In: Stevens SS, ed. *Handbook of experimental psychology.* New York: Wiley, 1951.
12 Baker B, Hardyck CD, Petrinovich LF. Weak measurement *v* strong statistics: an empirical critique of SS Stevens' proscriptions on statistics. *Educational and Psychological Measurement* 1966;**26**:291–309.
13 Wright BD, Masters GN. *Rating scale analysis.* Chicago: MESA Press, 1982.
14 Cohen J, Cohen P. *Applied multiple regression: correlation analysis for the behavioral sciences.* Hillsdale, New Jersey: Lawrence Erlbaum, 1983.
15 Linacre JM, Wright BD. *Winsteps/bigsteps Rasch-model computer program. Version 2.85.* Chicago: MESA Press, 1998.
16 Stilwell J, Hawley C, Stilwell P, *et al. The national traumatic brain injury study: report.* University of Warwick, Coventry: Centre for Health Services Studies, 1998.
17 Spector PE. Summated rating scale construction: an introduction. In: Lewis-Beck MS, ed. *Basic measurement.* London: Sage Publications, 1994.
18 SPSS. SPSS for Windows Version 7.5.1. Chicago: SPSS, 1996.
19 Dickson HG, Köhler F. The multi-dimensionality of the FIM motor items precludes an interval scaling using Rasch analysis. *Scand J Rehabil Med* 1996;**26**:159–62.