

VALIDATION OF THE EARLY FUNCTIONAL ABILITIES SCALE: AN ASSESSMENT OF FOUR DIMENSIONS IN EARLY RECOVERY AFTER TRAUMATIC BRAIN INJURY

Ingrid POULSEN, PhD^{1,3}, Svend KREINER, Cand. stat² and Aase W. ENGBERG, DMSci¹

From the ¹Department of Neurorehabilitation, Traumatic Brain Injury, Rigshospitalet, Copenhagen, ²Section of Biostatistics, Copenhagen University and ³Section of Nursing Science, Health, Aarhus University, Aarhus, Denmark

Objective: The Early Functional Abilities scale assesses the restoration of brain function after brain injury, based on 4 dimensions. The primary objective of this study was to evaluate the validity, objectivity, reliability and measurement precision of the Early Functional Abilities scale by Rasch model item analysis. A secondary objective was to examine the relationship between the Early Functional Abilities scale and the Functional Independence Measurement™, in order to establish the criterion validity of the Early Functional Abilities scale and to compare the sensitivity of measurements using the 2 instruments.

Methods: The Rasch analysis was based on the assessment of 408 adult patients at admission to sub-acute rehabilitation in Copenhagen, Denmark after traumatic brain injury.

Results: The Early Functional Abilities scale provides valid and objective measurement of vegetative (autonomic), facio-oral, sensorimotor and communicative/cognitive functions. Removal of one item from the sensorimotor scale confirmed unidimensionality for each of the 4 subscales, but not for the entire scale. The Early Functional Abilities subscales are sensitive to differences between patients in ranges in which the Functional Independence Measurement™ has a floor effect.

Conclusion: The Early Functional Abilities scale assesses the early recovery of important aspects of brain function after traumatic brain injury, but is not unidimensional. We recommend removal of the “standing” item and calculation of summary subscales for the separate dimensions.

Key words: construct validity; loglinear Rasch model; Rasch model; rehabilitation; traumatic brain injury; validation; Early Functional Abilities.

Accepted Oct 30, 2017; Epub ahead of print Jan 9, 2018

J Rehabil Med 2018; 50: 00–00

Correspondence address: Ingrid Poulsen, Department of Neurorehabilitation, Traumatic Brain Injury, Rigshospitalet, Kettegaard Allé 30, DK 2650, Hvidovre, Denmark. E-mail: Ingrid.Poulsen@regionh.dk

Restoration of brain function after severe traumatic brain injury (TBI) is a multi-faceted process that can be appraised by means of valid, objective and reliable measurement scales. The Early Functional Abilities (EFA) scale is one such instrument (1). It is used in German-speaking countries (1), Norway (2) and Denmark (3), and was developed for the assess-

Table I. Early Functional Abilities (EFA) scale items

EFA scale items
<i>Vegetative functions</i>
1. Stability of vegetative functions
2. Wakefulness
3. Tolerance for being in a supine position
4. Excretory functions (bladder and bowel)
<i>Facio-oral functions</i>
5. Oral stimulation and hygiene
6. Swallowing function
7. Tongue movements and chewing
8. Facial expression
<i>Sensorimotor functions</i>
9. Tone adaptation
10. Head control
11. Trunk control in sitting
12. Transfer from sitting to sitting
13. Standing position
14. Voluntary movements
15. Locomotion/mobility in wheelchair
<i>Perceptual and cognitive functions</i>
16. Tactile information
17. Visual information
18. Auditory information
19. Communication
20. Problem solving in activities of daily living

ment of functional recovery in the early stage after brain injury (1). The EFA comprises 20 items (Table I), on which the functional ability of the patient is scored using a 5-point scale (Table II).

The EFA is not the only instrument that attempts to measure aspects of restoration of brain function, but, compared with other generally used scales, the EFA aims to measure more aspects and to be more sensitive at lower levels of functioning.

The Functional Independence Measurement™ (FIM™) (4) and the Coma Recovery Scale-Revised (CRS-R) (5) are 2 such scales, measuring somewhat different aspects of restoration of brain function. The FIM™ is a measure of activities of daily living (ADL). Many studies have established the validity of the FIM™, but it suffers from a considerable floor effect. Thus, the FIM™ does not cover very low levels of functioning and may not be suitable for indicating early signs of recovery of function in ADL (6). The CRS-R (5), developed to assess the level of consciousness, is an elaborate instrument that has been shown to provide a unidimensional measure of the level of consciousness.

To our knowledge, only 4 studies (1–3, 7) have addressed issues relating to reliability and criterion validity of the EFA. Inter-rater reliability was examined and

Table II. The 5 Early Functional Abilities (EFA) levels, with responses rated on a scale from 0 to 4

0 point	1 point	2 points	3 points	4 points
Lack of function; no function observed	A trace of the function is present, but unstable and not specific or purposeful; very reduced function	Obviously present; stable specific, reproducible and purposeful, but not subtle; moderately reduced function	Stable, specific, purposeful and nuanced; slightly reduced function	Practically fully functioning; nearly normal functioning

shown to range from good to excellent (1, 2), while concurrent criterion validity was confirmed by the very high correlation ($r=0.86$) between the EFA and the FIMTM by Heck et al. (1). The correlation between the EFA and the FIMTM was also found by Stubbs et al. (3), who showed that when FIMTM was 18 (i.e. the lowest possible score), it was possible to observe improvement in function in 21% of patients admitted to brain injury rehabilitation. Thus, the EFA compensates for the floor effect of the FIMTM. However, the same study found that EFA had a ceiling effect also found by Heck et al. (1) who therefore recommended that at total EFA scores above 70, assessment by FIMTM should be preferred. Hankemeier & Rollnik (7) studied the concurrent and prognostic criterion validity of the EFA scale and found that it correlates with morbidity, length of stay, ADL and outcome.

Since none of the above studies have addressed the more challenging issues of validity and objectivity, the primary aim of the current study was to evaluate the validity, objectivity, reliability and measurement precision by assessing the degree to which the EFA items fit one or more Rasch models, because fit to Rasch models implies that measurement is both valid and objective. In this respect, it is important to note that the use of a total EFA score implicitly assumes that measurement is unidimensional, with one latent trait lying behind responses to all 20 items. As seen in Table I, the 20 items are divided into 4 subsets, defining 4 different subscales, as follows: *i*: vegetative (autonomic) function (VF) subscale; *ii*: facio-oral function (FOF) subscale; *iii*: sensorimotor function (SMF) subscale; *iv*: perceptual & cognitive function and ADL (PCF) subscale.

The 4 functions refer to qualitatively different aspects of brain functioning. Stabilization of vegetative functions is an important prerequisite for the readiness of the patient to engage in systematic rehabilitation, while facio-oral functions, including swallowing and facial expression, are vital for survival and social acceptability. Sensorimotor functions are essential to mobility, just as perceptual cognitive functions and mastering of ADL are decisive for interaction with other people and society in general. Thus, the analysis of validity and objectivity will be a 2-step process. In the first step, the fit of items to Rasch models will be assessed for each of the 4 subscales, after which the second step will test whether the subscales measure the same or different latent traits.

Based on this analysis, and given the fit to the Rasch models is adequate, the secondary aim of this study is to analyse the correlation between the EFA subscales and the FIMTM, in order to assess the degree of criterion validity of the EFA and compare the sensitivity of measurements on the 2 instruments.

METHODS

The Rasch analysis included item responses from 408 adult patients with severe TBI admitted for sub-acute neuro-rehabilitation to the Department of Neurorehabilitation, Rigshospitalet, Denmark, between October 2000 and December 2010. A subsample of 49 patients was also scored by FIMTM for analyses comparing the EFA with the FIMTM. Engberg et al. (8) describe the original admission criteria. Located in an acute hospital, the department was established to provide early intensive inter-disciplinary neuro-rehabilitation for the most severely brain-injured patients.

During rehabilitation EFA data were collected at admission, every 2 weeks, and at discharge. For the analysis in this paper, however, only the admissions dataset was used. Previous studies of EFA scored item responses from 1 to 5. However, during our analysis response categories were scored from 0 to 4, so that the total EFA score ranged from 0 to 80 points, where 80 indicates that a patient has no substantial reduction in functional ability. The decision to re-score responses in this way was based on the facts that psychometric models always score items in this way, and that the choice between scoring 1–5 or 0–4 has no implications for validity and reliability.

When the EFA scale was implemented in Denmark, we translated the original German version into Danish and back-translated it by standardized methods (9). Data handling was approved by the Danish Data Protection Agency.

Statistical methods

The sample of patients was described in terms of number and frequency for categorical data and mean, standard deviation (SD), and median for continuous variables.

Analysis of construct validity and objectivity of the EFA was performed by item analyses using Rasch models and graphical loglinear Rasch models. Details of these analyses are given below and in the Appendix S1[†].

Finally, criterion validity was assessed by the correlation between the EFA and the FIMTM for the subset of patients for whom both were available. Since the EFA and FIMTM scores are ordinal scores, Kendall's non-parametric measure of correlation (tau) was used to measure the strength of the association between them.

[†]<http://www.medicaljournals.se/jrm/content/?doi=10.2340/16501977-2300>

The Rasch model

Items fitting a Rasch model exhibit a number of properties that psychometricians refer to as internal construct validity or criterion-related construct validity (10): (i) unidimensionality; (ii) monotonicity, in the sense that expected item scores are increasing functions of the values of the latent variable; (iii) local independence; and (iv) no differential item functioning (DIF).

Several psychometric models satisfy the requirements of criterion-related construct validity, but the Rasch model is unique among these because it is the only model with a statistically sufficient person score in which measurement is objective in Rasch's sense of the word.

Graphical loglinear Rasch models

Health-related scales, such as the EFA scale, rarely contain items that satisfy all requirements of local independence and no DIF. In such cases graphical loglinear Rasch models (GLLRMs) (11) may be an option. In GLLRMs, items are permitted to be locally dependent and function differentially in different groups, but local dependence and DIF must be uniform in the sense that the strength of the association between items and between items and exogenous covariates is the same for all persons, irrespective of the value of the person parameter (11).

Two consequences follow from the assumption of uniform dependence and DIF. First, uniform DIF implies that items do fit the Rasch models in the different groups defined by the covariates, but that item parameters differ between groups. Secondly, uniform local dependence implies that super-items defined by the sum of the dependent items will fit the Rasch model. This cannot happen if local dependence is non-uniform.

One way to deal with an item with uniform DIF relative to a covariate is to treat it as a set of different items that have been administered to different groups. Such procedures are sometimes referred to as item splitting. From this, it follows that item splitting and calculation of super-items, result in a modified set of items that do fit ordinary Rasch models. Since the summary raw score over the modified items is the same as the raw score from the ordinary Rasch model, and since this raw score also is sufficient for the person parameter in the GLLRM, Kreiner & Christensen (11) claim that measurement by GLLRMs is essentially valid and objective, even though the original items violate standard requirements of both validity and objectivity.

Item analysis by Rasch models and graphical loglinear Rasch models

The purpose of our Rasch analysis of EFA was two-fold. The first part of the analysis tested the validity and objectivity of EFA measurement by tests of fit of items to the models. The second part assessed the quality of measurement by analysis of measurement errors, targeting and reliability.

During the analysis of fit of items to Rasch models, the assumptions are tested by:

- overall tests that item properties are homogeneous across all levels of the latent trait underlying responses to items;
- overall tests that items do not function differentially (no DIF) relative to Sex and Age;
- tests of local independence for all pairs of items;
- tests of no DIF for separate items relative to Sex and Age;
- item fit statistics comparing observed responses to items with the responses expected by the Rasch model;
- tests of unidimensionality.

A recent paper by Lundgren Nilsson & Tennant (12) describes issues in modern Rasch analysis, and a paper by Christensen et al. (13) sets out the technical details. The analysis used fit statistics that do not depend on the distribution of persons, and with known asymptotic distributions, thus the results can also be trusted in large sample situations. Andersen's (14) conditional likelihood ratio tests were used for the overall tests of homogeneity, Kelderman's conditional likelihood ratio test was used for tests of local dependence, and tests of no DIF were used for specific pairs of items and covariates (15).

Item fit statistics include conditional infits and outfits and tests of differences between observed and expected correlations of items and rest-scores, as suggested by Christensen & Kreiner (16). Infits and outfits provide 2 different ways to summarize residuals measuring differences between observed and expected responses. Both fits are standardized so that fit statistics are equal to 1 if the fit is perfect. In addition to these fit statistics, we also compared the observed and expected correlations between items and rest scores, where the rest score for an item is equal to the total score for all other items.

Measurement

Measurement works in the same way in GLLRMs as in Rasch models. The total score is a sufficient statistic, and estimates of person parameters provide measurement on interval scales. As in Rasch models, weighted maximum likelihood estimates (17), which are known to have less bias than other estimates of person parameters in Rasch models, may also be used.

Quality of measurement by GLLRMs is assessed in 3 different ways. First, by calculation of standard errors of the estimates of the person parameters, referred to as standard errors of measurement (SEM). Secondly, by assessment of the degree to which the items and the EFA scores target the study population. Finally, calculations of reliability describe the degree to which the EFA subscales are able to distinguish between persons according to their function.

Estimation of person parameters converts total raw scores to measures on interval scales. Tables of these estimates are included in the Appendix SI¹ together with the standard errors of the estimates, often referred to as SEM, during Rasch analyses. We are of the opinion that the estimate of the person parameter is the optimal measure, but we are aware that many users of Rasch models prefer to use the total raw score as a measure. In such cases, it is important to recall that the total score also has an SEM. These are also calculated and reported in the Appendix SI¹.

The SEM by Rasch models describes the precision of measurement at an individual level. Since SEM depends on the person parameter of the model, it follows that measurement will be more precise for certain persons than for others. This raises questions of the degree to which EFA is appropriate for the patient population. This issue is addressed in 2 different ways; first, by analysis of the degree to which EFA targets the study population; and, secondly, by calculation of reliability.

Targeting. Targeting is, in most cases, assessed in an informal way by so-called item maps comparing the distribution of the item thresholds of the Rasch model with the distribution of the persons, and requiring that the distribution of thresholds cover the distribution without too many thresholds lying either far below or far above the persons. Such maps are included in the Appendix SI¹.

The *target* of a scale is the value of the person parameter where SEM is minimized, and the *true score at target* is the

expected raw score of persons for persons at target. To assess the degree to which the score targets the population we estimated the mean of the person parameter and compared the mean SEM of the population with the SEM at target. If there was a large difference between the target and the population mean and if the mean SEM in the population was much larger than the SEM at target, we concluded that the scale was out of target relative to the study population.

Reliability. Psychometrics defines reliability as the ratio between variance of the true score in the population and the variance of the observed scores. Reliability depends on both the SEM and the true variance of the population. It is therefore important to emphasize that measurement precision measured by SEM and reliability are 2 different concepts. Reliability is a measure of the degree to which a scale is able to separate the persons in the study population, and Rasch analyses often describes measures of reliability as indices of person separation. If measurements are used to track the development of single patients, it is only the SEM that counts, and reliability is irrelevant. Assessment of change over time will have little power if SEMs are large, and have more power if SEMs are ignorable whether or not the reliability appears to be adequate. If SEMs are ignorable, the power will be high, even though reliability is poor.

Criterion validity of the Early Functional Abilities scale. Finally, to examine the criterion validity of the EFA, the association between the FIM™ and the EFA for the 49 patients with available FIM™ information was analysed.

Assessment of significance

Significance was evaluated at a 5% critical level after adjustment for multiple testing by the Benjamini & Hochberg procedure (18). We distinguish between weak to moderate evidence against the model, where *p*-values are larger than 0.01, and stronger evidence when *p*-values are less than 0.01. Weak evidence against the Rasch model provided by the overall conditional likelihood ratio (CLR) tests is not regarded as conclusive unless it is supported by evidence against the fit of items or evidence of either local dependence or DIF for specific items.

Computer programs

DIGRAM (19, 20) was used for the item analysis by Rasch models and GLLRMs. SPSS was used for descriptive analyses of data and for calculation of Kendall's tau.

RESULTS

This section summarizes and discusses the most important results of the item analyses of the 4 EFA subscales. Additional results, including estimates of the item and person parameters, and additional comments on the analyses are provided in the Appendix SI¹.

Characteristics of the patients

Table III describes the sample of patients. The subsample of 49 patients (not shown here) did not differ significantly from the total sample.

Table III. Characteristics of the study population

Variable	<i>n</i>	Frequency, %	Mean	Median	SD
Sex	408				
Male	310	76.0			
Female	98	24.0			
Age, years ^a	408		42.7	41	18.7
GCS ^a	405		10.9	11	2.9
3–9	128	31.6			
10–12	121	29.9			
13–14	140	34.5			
15	16	4.0			
FIM™ ^a	49		35.4	18	27.0
FIM™=18	26	53.3			
ISS ^a	403		30.9	29	9.4
0–24	36	8.4			
25+ (severe injury)	367	91.6			
PTA (days) ^b	397		105.7	53	105.1
GOS ^c	397				
Dead	16	3.9			
Persistent vegetative state	11	2.7			
Severe disability	217	53.4			
Moderate disability	127	31.3			
Good recovery	35	8.6			

^aAt admission to sub-acute rehabilitation; ^bAt discharge or follow up; ^cAt discharge.

SD: standard deviation; GCS: Glasgow Coma Scale (26); FIM™: Functional Independence Measure™ (4); ISS: Injury Severity Score (27); PTA: post-traumatic amnesia (28); GOS: Glasgow Outcome Scale (29).

Analyses of fit of items to Rasch models and graphical loglinear Rasch models

Fit to Rasch models was rejected for all 4 EFA subscales. Subsequent analysis by GLLRMs showed that the reason was due to local dependence, but also found no evidence of DIF. Table IV shows the pairs of items that had to be included in GLLRMs to obtain fit of items to data, together with CLR tests supporting the claims of local dependence.

The local dependencies in Table IV define 4 different GLLRMs. In addition to adding local dependence to the models, it was necessary to eliminate Item 13

Table IV. Conditional likelihood ratio^a tests of local independence

	χ^2	Df	<i>p</i> -value
Vegetative functions			
Stability (1) & Tolerance (3)	52.5	16	<0.00005
Tolerance (3) & Excretory functions (4)	43.9	16	0.0002
Facial-oral functions			
Oral stimulation (5) & Tongue movements (7)	58.5	16	<0.00005
Swallowing (6) & Tongue movements (7)	107.4	16	<0.00005
Sensorimotor functions			
Tone adaption (9) & Voluntary movements (14)	41.9	16	0.0004
Head Control (10) & Trunk control in sitting (11)	35.7	16	0.0032
Head Control (10) & Mobility in wheelchair (15)	38.4	16	0.0013
Transfer from sitting (12) & Mobility in wheelchair (15)	39.6	16	0.0009
Perceptual and cognitive function			
Tactile information (16) & Auditory information (18)	39.1	16	0.0011

^aConditional likelihood ratio tests are asymptotically χ^2 distributed. All tests are calculated under the graphical loglinear Rasch models (GLLRMs) defined by the dependencies presented in the model. The numbers in parentheses refer to item numbers of the Early Functional Abilities (EFA) scale.

Table V. Overall conditional likelihood ratio tests of homogeneity^a and no differential item functioning (DIF)

Functions	Model	Homogeneity	DIF relative to sex	DIF relative to age
Vegetative	Rasch	$\chi^2 = 44.4$, df = 15, $p < 0.0005$	$\chi^2 = 11.7$, df = 15, $p = 0.70$	$\chi^2 = 25.5$, df = 30, $p = 0.70$
	GLLRM	$\chi^2 = 9.9$, df = 45, $p = 1.00$	$\chi^2 = 45.0$, df = 45, $p = 0.47$	$\chi^2 = 90.8$, df = 90, $p = 0.46$
Facio-oral	Rasch	$\chi^2 = 62.0$, df = 15, $p < 0.0005$	$\chi^2 = 27.8$, df = 15, $p = 0.023$	$\chi^2 = 47.9$, df = 30, $p = 0.020$
	GLLRM	$\chi^2 = 15.4$, df = 37, $p = 1.00$	$\chi^2 = 59.6$, df = 37, $p = 0.011$	$\chi^2 = 95.3$, df = 74, $p = 0.049$
Sensorimotor	Rasch	$\chi^2 = 46.7$, df = 27, $p = 0.011$	$\chi^2 = 31.7$, df = 27, $p = 0.24$	$\chi^2 = 59.9$, df = 54, $p = 0.27$
	GLLRM ^b	$\chi^2 = 42.2$, df = 73, $p = 1.00$	$\chi^2 = 94.4$, df = 73, $p = 0.047$	$\chi^2 = 172.0$, df = 146, $p = 0.070$
Perceptual & cognitive	Rasch	$\chi^2 = 20.3$, df = 19, $p = 0.38$	$\chi^2 = 22.4$, df = 19, $p = 0.27$	$\chi^2 = 56.3$, df = 38, $p = 0.028$
	GLLRM	$\chi^2 = 23.3$, df = 33, $p = 0.89$	$\chi^2 = 41.1$, df = 33, $p = 0.16$	$\chi^2 = 80.3$, df = 66, $p = 0.11$

^aThe homogeneity test compares item parameters for individuals with high and low scores; ^bItem 13 excluded.
DIF: differential item functioning; GLLRM: graphical loglinear Rasch model.

(Standing) from the subscale measuring Sensorimotor functions. In addition, the fit to the GLLRMs was accepted as shown in Table V, which also shows the overall CLR tests rejecting fit to the Rasch model.

In addition to supporting the claim that local dependence has been taken care of by the GLLRMs, Table V also indicates that there is no DIF. Elaboration on this can be found in the Appendix SI¹.

Table VI with item fit statistics provides further support of the fit to the GLLRMs. The γ correlations are rank correlations for ordinal categorical data that are similar to Kendall's tau. All fit statistics comfortably accept the fit of the items to the models.

Targeting and reliability

Information on targeting and reliability can be found in the Appendix SI¹.

Reliability is high for all for EFA subscales, but the VF, FOF and SMF subscales are somewhat off-target; they target patients with a higher level of functioning than the study population.

Unidimensionality or multidimensionality

Following the successful fit of the 4 subscales to GLLRMs, we examined the degree to which the VF, FOF, SF and PCF subscales measure the same latent trait, in order to justify summarizing the different subscales into a single overall EFA scale. Table VII compares the observed correlations among subscales with the correlations expected by the unidimensional model. Since the observed correlations in all cases are significantly weaker than expected by the unidimensional model, we conclude that the EFA subscales measure 4 distinct and qualitatively different functions

Table VI. Item fit statistics calculated under graphical loglinear Rasch models with local dependence among items. Since items and sub-scores are measured on ordinal scales, we measure the correlations between items and rest scores by the γ coefficient proposed by Goodman & Kruskal (30)

Items	Outfit	p -value	Infit	p -value	Observed γ	Expected γ	p -value
Vegetative							
1. Stability	1.00	1.00	1.029	0.73	0.58	0.59	0.85
2. Wakefulness	0.86	0.63	0.94	0.55	0.70	0.71	0.85
3. Tolerance	0.94	0.72	0.95	0.61	0.50	0.49	0.91
4. Excretory functions	1.05	0.78	1.17	0.11	0.67	0.67	0.99
Facio-oral							
5. Oral stimulation	0.90	0.28	0.94	0.58	0.72	0.71	0.72
6. Swallowing function	1.05	0.57	1.07	0.40	0.79	0.80	0.61
7. Tongue movements	0.86	0.84	1.01	0.96	0.85	0.84	0.63
8. Facial expression	0.93	0.40	0.98	0.84	0.71	0.70	0.68
Sensorimotor							
9. Tone adaption	1.03	0.69	1.01	0.83	0.79	0.79	0.97
10. Head control	0.95	0.59	0.97	0.73	0.79	0.78	0.79
11. Trunk control	1.05	0.11	0.96	0.58	0.79	0.78	0.65
12. Transfer from sitting	0.72	0.67	0.97	0.83	0.88	0.87	0.43
14. Voluntary movements	1.04	0.76	1.17	0.045	0.77	0.80	0.22
15. Locomotion/Mobility in wheelchair	0.81	0.31	0.97	0.81	0.82	0.80	0.50
Perceptual & cognitive							
16. Tactile information	1.10	0.35	1.12	0.14	0.78	0.80	0.35
17. Visual information	1.13	0.27	1.16	0.047	0.83	0.86	0.12
18. Auditory information	0.80	0.036	0.83	0.042	0.85	0.81	0.06
19. Communication	1.17	0.16	1.10	0.22	0.85	0.86	0.62
20. Problem solving in ADL	0.69	0.25	0.84	0.07	0.89	0.86	0.08

ADL: activities of daily living.

Table VII. Analysis of multidimensionality by assessment of the correlation among sub-scores. Since sub-scores are measured on ordinal scales, we measured the correlation using the γ coefficient proposed by Goodman & Kruskal (30)

Subscales	Expected correlation	Observed correlation	<i>p</i> -value
Vegetative, facio-oral	0.692	0.643	0.0000
Vegetative, sensorimotor	0.758	0.656	0.0000
Vegetative, perceptual & cognitive	0.763	0.605	0.0000
Facio-oral, sensorimotor	0.782	0.697	0.0000
Facio-oral, perceptual & cognitive	0.809	0.747	0.0000
Sensorimotor, perceptual & cognitive	0.823	0.735	0.0000

and that the outcomes on the subscales therefore should not be summarized into a single EFA score.

Criterion validity – Kendall's tau correlation of the EFA and the FIMTM

Table VIII shows the correlation between EFA subscales and FIMTM subscales measuring motor and cognitive functions, together with the total FIMTM scale. Table VIII confirms criterion validity. It is noteworthy that there are virtually no differences between the strengths of the correlation between the different EFA subscales, on the one hand, and the FIMTM scores on the other. This may suggest that the summary of ADL activities registered by FIMTM depends on an unclear mixture of the 4 functions measured by the EFA.

Fig. 1 shows that the EFA scale is able to discriminate between 53% of the patients, whereas the FIMTM cannot register any change in ADL because of a “floor” of low scores (FIMTM=18).

DISCUSSION

One of the main findings of the present study is that the EFA subscales measure features of early recovery after severe TBI. Our analysis supports our conclusions, since the subscales, which summarize responses to items on the VF, FOF, SF and PCF subscales, fulfil the requirements of GLLRMS. Item analysis, however, showed that item fit statistics disclose a strong misfit for Item 13 (standing), suggesting that it should be removed from the EFA scale. It also follows that the

Table VIII. Correlation between the Early Functional Abilities (EFA) subscales and the Functional Independence MeasurementTM (FIMTM)

EFA subscales	FIM TM		
	Motor function	Cognitive function	Total
VF	0.65	0.67	0.65
FOF	0.70	0.67	0.68
SMF	0.66	0.64	0.65
PCF	0.63	0.68	0.67

Since subscales are measured on ordinal scales, we measured the correlation using Kendall's tau. All correlations are highly significant, with $p < 0.0005$. VF: vegetative (autonomic) function subscale; FOF: facio-oral function subscale; SMF: sensorimotor function subscale; PCF: perceptual & cognitive function and ADL subscale.

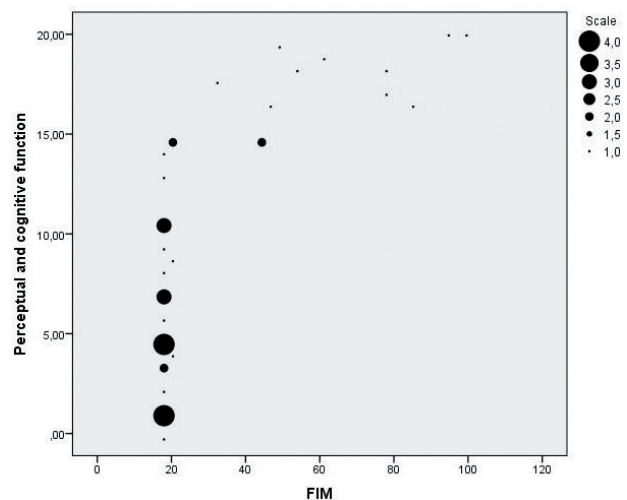


Fig. 1. Association between the PCF subscale and the total FIMTM score.

summated scores for each of these 4 subscales can be transformed into values on interval scales, where a unit of measurement is the logit, if this is considered convenient (21). Furthermore, the analysis of criterion validity relative to FIMTM confirmed EFA's validity, showing that low EFA scores have a much wider range of sensitivity than FIMTM. In rehabilitation research into the sub-acute stage of patients with severe TBI a well-known problem is describing functional changes over time. Hart et al. (6), for example, examine functional outcome by means of the FIMTM in patients admitted to sub-acute rehabilitation after severe TBI. Approximately 10% of the patients with the most severe TBI were excluded from analyses due to a FIMTM score of 18 during their rehabilitation. Consequently, we suggest using the EFA as a supplement to the FIMTM when studying changes over time in this patient group.

If the EFA subscales had proved to measure the same underlying latent phenomena then they could conveniently be combined into a total EFA score, but the analysis of unidimensionality rejected this, showing that a total EFA score does not provide a statistically sufficient description of the patient's functional state. The fact that, according to our analysis, the sub-scores of the 4 subscales should not be summed reflects the complicated nature of restoration of brain function after TBI. Nevertheless, Table VII shows that the subscales are correlated with each other.

Compared with the CRS-R (5), the EFA scale assesses different aspects of restoration, pointing more clearly to areas where rehabilitation efforts are particularly needed. Thus, the EFA reflects the clinical reality in a much broader sense than the CRS-R. However, the CRS-R, assessing only the level of consciousness, has the benefit of being unidimensional, making it suitable for research and practice, especially in patients

in the vegetative/unresponsive wakefulness state and minimally conscious state (5). The disadvantage of including several different aspects is that neither the FIM™ (22, 23) nor the EFA is unidimensional. Cook et al. (24) recognized this as long ago as 1994, when they proposed a profile based on assessment of each of the 18 FIM™ items. We show how 4 subscale-total scores constitute a single adequate EFA profile. Therefore, assessment of recovery after TBI should use profiles containing sub-scores, or, even better, logits in which sub-scores are converted to proper measurements on interval scales, to avoid the difficulties with ordinal scales pointed out by Merbitz et al. (25).

Methodological issues

According to Lundgren Nilsson & Tennant (12) the deletion of items from a scale should be considered a last resort for a variety of reasons. This is particular important for health-related scales that often contain very few items. Instead, they suggest a number of ways to avoid the problems caused by misfit to the Rasch model that is equivalent to item analysis by GLLRMs without the formal framework. The GLLRMs split items with DIF and treat them as different items that are unobserved in some groups. Instead of estimating item parameters in each of the groups, a GLLRM analysis selects one group as the reference group and estimates differences or contrasts between the item parameters in the reference group and the item parameters in other groups, and refers to these contrasts as inter-action parameters. Instead of calculating super-items, as suggested by Nilsson & Tennant, a GLLRM analysis estimates interaction parameters between items and uses these parameters to estimate the item parameters of the super-items. In the end, there is little difference between the results of Lundgren Nilsson & Tennant's analysis of DIF and local dependence and the results of a GLLRM analysis, but the GLLRM analysis is advantageous for several reasons. First, the estimates of the interaction parameters make it easier to understand and interpret the association between dependent items and the effects of covariates causing DIF. Secondly, during GLLRM analysis it is also possible to assess the fit of single items in cases where the misfitting item is 1 of a set of dependent items. A test of fit of super-items with 1 misfitting item will either show that the super-items do not fit or fail to indicate that there is a problem. Finally, GLLRMs may also be used if all items are directly or indirectly dependent.

Study limitations

A limitation of this study is that it includes patients at admission to sub-acute rehabilitation with severe TBI

from only one clinical setting. However, since these patients are rehabilitated in only 2 nationwide centres in Denmark, it is likely that our sample consists of an unselected patient group. Thus, we believe that our results are representative of patients with TBI. As the EFA is used in children too, there is a need for further studies examining its validity in children with TBI.

In conclusion, the EFA subscales provide valid, objective and reliable measures of early functional abilities that can be useful to assess the degree of recovery after TBI. After removal of the "standing item" from the SMF subscale, the 4 subscales provide statistically sufficient measures of qualitatively different functions that can be converted to measurements on interval scales by estimates of the person parameters of the GLLRMs. Since measurement is clearly multidimensional, we cannot recommend that the 4 EFA subscales are summarized into a total EFA scale because this score does not provide a sufficient description of the functional ability and cannot be converted to an interval-scaled measure of functional ability.

Finally, this study also confirmed that measurement with the EFA scale is criterion valid, and that it is able to discriminate between patients with inadequate functional abilities, whereas the FIM™ cannot distinguish between patients.

REFERENCES

1. Heck G, Steiger-Bächler G, Schmidt T. Early Functional Abilities (EFA) – eine Skala zur Evaluation von Behandlungsverläufen in der neurologischen Frührehabilitation. *Neurol Rehabil* 2000; 6: 125–133.
2. Alvsaker K, Walther SM, Kleffegard I, Mongs M, Draegebo RA, Keller A. Inter-rater reliability of the early functional abilities scale. *J Rehabil Med* 2011; 43: 892–899.
3. Stubbs PW, Pallesen H, Pedersen AR, Nielsen JF. Using EFA and FIM rating scales could provide a more complete assessment of patients with acquired brain injury. *Disabil Rehabil* 2014; 36: 2278–2281.
4. Corrigan JD, Smith-Knapp K, Granger CV. Validity of the functional independence measure for persons with traumatic brain injury. *Arch Phys Med Rehabil* 1997; 78: 828–834.
5. Giacino JT, Kalmar K, Whyte J. The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. *Arch Phys Med Rehabil* 2004; 85: 2020–2029.
6. Hart T, Kozlowski AJ, Whyte J, Poulsen I, Kristensen K, Nordenbo A, et al. Functional recovery after severe traumatic brain injury: an individual growth curve approach. *Arch Phys Med Rehabil* 2014; 95: 2103–2110.
7. Hankemeier A, Rollnik JD. The Early Functional Abilities (EFA) scale to assess neurological and neurosurgical early rehabilitation patients. *BMC Neurol* 2015; 15: 207.
8. Engberg AW, Liebach A, Nordenbo A. Centralized rehabilitation after severe traumatic brain injury – a population-based study. *Acta Neurol Scand* 2006; 113: 178–184.
9. Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. *J Adv Nurs* 2004; 48: 175–186.
10. Rosenbaum PR. Criterion-related construct-validity. *Psychometrika* 1989; 54: 625–733.
11. Kreiner S, Christensen KB. Validity and objectivity in health related summated scales: analysis by graphical loglinear Rasch models. In: von Davier M, Carstensen, CH, editor.

- Multivariate and mixture distribution Rasch models – extensions and applications. New York: Springer Verlag; 2007, p. 329–346.
12. Lundgren Nilsson A, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM) revisited. *J Rehabil Med* 2011; 43: 884–891.
 13. Christensen KB, Kreiner S, Mesbah M. Rasch models in health. London: ISTE Ltd & Wiley & Sons; 2013.
 14. Andersen EB. Asymptotic properties of conditional maximum-likelihood estimators. *J R Stat Soc Series B Methodol* 1970; 32: 283–301.
 15. Kelderman H. Loglinear Rasch model tests. *Psychometrika* 1984; 49: 223–245.
 16. Christensen KB, Kreiner S. Item fit statistics In: Christensen KB, Kreiner S, Mesbah M, editors. Rasch models in health. London: ISTE Ltd & Wiley & Sons; 2013, p. 83–104.
 17. Warm TA. Weighted likelihood estimation of ability in item response theory. *Psychometrika* 1989; 54: 427–450.
 18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 1995; 57: 289–300.
 19. Kreiner S. Introduction to DIGRAM. Copenhagen: Department of Biostatistics, University of Copenhagen; 2003.
 20. Kreiner S, Nielsen T. Item analysis in DIGRAM: guided tours. Copenhagen: Department of Biostatistics, University of Copenhagen; 2006.
 21. Andrich D. Rasch models for measurement. London: Sage Publications; 1988.
 22. Nilsson ÅL, Sunnerhagen KS, Grimby G. Scoring alternatives for FIM in neurological disorders applying Rasch analysis. *Acta Neurol Scand* 2005; 111: 264–273.
 23. Yamada S, Liu M, Hase K, Tanaka N, Fujiwara T, Tsuji T, et al. Development of a short version of the motor FIM for use in long-term care settings. *J Rehabil Med* 2006; 38: 50–56.
 24. Cook L, Smith DS, Truman G. Using Functional Independence Measure profiles as an index of outcome in the rehabilitation of brain-injured patients. *Arch Phys Med Rehabil* 1994; 75: 390–393.
 25. Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. *Arch Phys Med Rehabil* 1989; 70: 308–312.
 26. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* 1974; 2: 81–84.
 27. Osler T, Rutledge R, Deis J, Bedrick E. ICISS: an international classification of disease-9 based injury severity score. *J Trauma* 1996; 41: 380–386; discussion 6–8.
 28. Zafonte RD, Mann NR, Millis SR, Black KL, Wood DL, Hammond F. Posttraumatic amnesia: its relation to functional outcome. *Arch Phys Med Rehabil* 1997; 78: 1103–1106.
 29. Pettigrew LE, Wilson JT, Teasdale GM. Assessing disability after head injury: improved use of the Glasgow Outcome Scale. *J Neurosurg* 1998; 89: 939–943.
 30. Goodman LA, Kruskal WH. Measures of association for cross classifications. *J Am Stat Assoc* 1954; 49: 732–764.