Supplementary material to article by I. Poulsen et al. "Validation of the early functional abilities scale: an assessment of four dimensions in early recovery after traumatic brain injury"

Appendix SI

This supplement provides additional information on methodology and on item analysis of the Early Functional Abilities (EFA) subscales.

Choice of fit statistics and assessment of significance

It follows from the sufficiency of the raw score in Rasch models and GLLRMs that item analysis can be performed without assumption of the distribution of the latent variable. In order to take advantage of this, the analysis in this study was based on the principles of conditional inference proposed and encouraged by Rasch (S1) and subsequently developed by Andersen (14, S2).

Many different fit statistics have been proposed for test of fits of items to item-response theory (IRT) and Rasch models and many of these have problems with correct assessment of significance in large sample studies because the distributions used for calculation of *p*-values do not apply as proper asymptotic distributions when the sample size increases towards infinity. In order to avoid these problems, we used conditional fit statistics where the asymptotic properties had been established, and parametric bootstrapping to estimate *p*-values of fit statistics where reliable asymptotic properties were unavailable. Our analysis assessed the overall fit of the model and the overall assessment of no DIF by Andersen's (14) CLR test. The fit of specific items to the Rasch model was assessed by conditional Infits and Outfits (16) and by comparison of the observed and expected correlation between scores for separate items and the rest-score over all other items, where the rest-score relative to a specific item is equal to the total score minus the score on the item (14, S3). Finally, the assumptions of local dependence and no DIF were tested by the CLR tests proposed by Kelderman (15) and by analyses of the partial association of items and exogenous covariates given a total score over other items, as described by Christensen & Kreiner (16).

Measurement quality

In order to assess the quality of measurement the following related issues had to be addressed:

- SEMs;
- the degree to which the measurement instrument targeted the patient population;
- measurement reliability.

We remind the reader that a measure of reliability is not a measure of the accuracy of measurement and that the use of Cronbach's alpha to provide a lower bound of reliability assumes that items are locally independent, which is not the case in the EFA. Cronbach's alpha was therefore not used during our analysis.

Classical test theory (CTT) defines reliability in 2 different ways. First, as the correlation between to equivalent forms of a test (S4) and second, as the ratio between the variances of the true and observed scores. CTT assumes that SEM does not depend on the true score and therefore defines reliability as:

$$r = \frac{\text{var(TS)}}{\text{var(TS)+SEM}^2}$$

where TS is the true score and Var(TS)+SEM^2 is the variance of the observed score. Under the assumptions of CTT this is equal to the correlation between equivalent forms of an instrument.

The second definition clearly shows that reliability *depends* on the SEM, it is also true that it depends as much on the variance of the true score in the study population. What reliability tells us is therefore not the degree to which measurement is precise, but the degree to which the measurement instrument is able to separate the persons in the study parameter in the correct way. Another measure of reliability would be an estimate of the probability that the measurements separate 2 randomly selected persons from the study population in the correct way. Such estimates are included in the section on measurement quality in this appendix.

Since the EFA items are locally dependent, we cannot use Cronbach's alpha for calculation of reliability. Instead, we use Hamond & Mesbah's Monte Carlo estimates (S5). This procedure works as follows:

- the first step estimates the mean and the variance of the person parameters in the study population;
- the second step generates a Monte Carlo (MC) sample of random person parameters from a normal distribution with the estimated mean and variance. During our analyses the sample size of the MC sample was equal to 10,000;
- the third step generates repeated *independent* scores for each person in the MC sample from the conditional distribution of the score given the person parameter;
- finally, Hammond & Mesbah (S5) estimated the reliability using the correlation of the repeated measurement in the MC sample, while we used the same MC sample, to also estimate the probability that a random person with a higher true score than another random person also has the highest observed score.

Finally, we note that it is common to define the reliability of measurement by Rasch models, where the true score is replaced by the true person estimate and the observed score by the estimate of the person parameter, and to rename the reliability of the reliability coefficient as person separation index (PSI), which is a much more meaningful term than CTT's "reliability". We did not calculate the PSI, because we know that reliability, as defined by CTT, in most cases is close to the PSI. Instead we calculated the probability of correct person separation, which is the same for scores and person estimates, and which is often very close to reliability and PSI.

Early Functional Abilities models

The EFA subscales are characterized by local dependence. Table IV, in the original paper, lists the pairs of items that have been found to be dependent, together with the CLR tests supporting the claims of dependence.

Since the items fit the GLLRMs where uniform dependencies are allowed, we conclude that we have no evidence against uniform local dependencies and interpret the evidence of dependence as evidence of local *response* dependence.

Estimates of item parameters

Subsets of GLLRM items that are directly or indirectly dependent define item components where super-items, defined by the sum of scores in components, are distributed as partial credit items. STable I shows the estimates of the item thresholds of the items and super-items defined by the models.

The VF model contains one item component defined by a chain of 3 items (Item 1 ↔ Item 3 ↔ Item 4), and where only Item 2 is locally independent of the other VF items. A few of the thresholds of the super-items are disordered, but this is not important, because it is known that local dependence may create super-items with disordered thresholds. From this point of view, it may be of greater concern that the thresholds of Item 2 are disordered.

The FOF scale also has one locally independent item and a chain of dependent items (Item 5 ↔ Item 7 ↔ Item 6). In this case, there is little evidence of disorder among item thresholds.

STable I. Item thresholds

	Item thresholds
Vegetative functions	
1+3+4. Stability + Tolerance + Excretory functions	−4.70, −2.68, −1.46, −0.89, −0.32, 0.07, 0.43, 0.60, 0.87, 1.15, 0.28, 0.96
2. Wakefulness	0.35, 1.98, −0.58, 3.94
Facial-oral functions	
5+6+7. Oral stimulation + Swallowing + Tongue movements	−3.97, −2.49, −1.24, −0.44, −0.49, −0.24, 0.56, 0.69, 0.78, 1.09,1.48, 1.49
8. Facial expression	−1.78, 0.64, 0.66, 3.03
Sensorimotor functions	
9+14. Tone adaptation + Voluntary movements	−2.62, −2.35, −0.91, −0.75, −0.25, 0.57, 0.15, 1.14
10+11+12+15 Head control + Trunk control in sitting + Transfer from sitting + Mobility in wheelchair	−2.43 −2.13 −1.33 −0.89 −0.52 −0.09 0.05 0.35 0.41 0.46 0.40 0.53 0.52 cont. 1.15 1.46 3.36
Perceptual and cognitive functions	
16+18. Tactile information + Auditory information	−6.33, −3.62, −2.61, −1.47, −0.65, 0.81, 2.12, 4.12
17. Visual information	−3.35, −1.28, 1.04, 3.03
19. Communication	−3,18, −1.17, 1.74, 3.20
20. Problem-solving in activities of daily living (ADL)	−1.26, −0.44, 2.87, 6.44

STable II. Comparison of observed and expected correlations among super-items and rest scores

Super-item	Rest score	Observed	Expected	*p*-value
Vegetative				
Items 1+3+4	Item 2	0.70	0.71	0.85
Facial-oral functions				
Items 5+6+7	Item 8	0.71	0.70	0.68
Sensorimotor				
Items 9+14	Items 10+11+12+15	0.73	0.75	0.36
Items 10+11+12+15	Items 9+14	0.73	0.75	0.36
Perceptual & cognitive				
Items 16+18	Items 17+19+20	0.87	0.86	0.35

The item structure of the SMF subscale is complicated, with 2 components and no independent items. The first component connects Items 9 and 14. The second component is defined by chain of 4 items (Item 11 ↔ Item 10 ↔ Item 15 ↔ Item 12). Thresholds are not required to be disordered in super-items, but there is little evidence of disorder found in the SMF subscale.

Finally, the PCF subscale has one component connecting Items 16 and 18, and 3 independent items. The PCF subscale is therefor closer to a pure Rasch scale than the other subscales. There is no evidence of disorder for the 3 independent items and very little evidence for the super-items.

Overall tests of fit

Table V, in the original paper, shows the overall tests of fit. There is weak evidence of DIF relative to Sex and Age, and these are not supported by the tests of DIF for separate items.

Item fit statistics

Table VI, in the original paper, shows the item fit statistics. The fit to the model is comfortably accepted by all the item fits.

For completeness, STable II compares the observed and expected correlations between super-items and rest scores without the super-item. These tests also support the fit of items to the GLLRM.

Test of local independence

According to the GLLRMs, certain pairs of items are locally dependent, whereas other pairs are locally independent; therefore, 2 tables of tests of local independence were needed to support the models.

Table IV, in the original paper, shows the test supporting claims of local dependence. STable III shows the tests supporting claims of local independence. Since the *p*-value has to be less than 0.002 after adjustment for multiple testing by the Benjamini-Hochberg procedure (18), we conclude that there is no significant

STable III. Conditional likelihood ratio tests of local independence

Items	χ^2	Df	*p*-value
Vegetative			
1 & 2	22.0	16	0.14
1 & 4	24.6	16	0.078
2 & 3	17.4	16	0.36
2 & 4	25.6	16	0.060
Facial-oral functions			
5 & 6	11.1	16	0.80
5 & 8	25.4	16	0.063
6 & 8	28.6	16	0.027
7 & 8	23.6	16	0.098
Sensorimotor			
9 & 10	30.6	16	0.015
9 & 11	1.8	16	1.00
9 & 12	15.2	16	0.51
9 & 15	16.4	16	0.42
10 & 12	14.2	16	0.58
10 & 14	14.9	16	0.53
11 & 12	32.7	16	0.008
11 & 14	18.1	16	0.32
11 & 15	16.4	16	0.43
12 & 14	10.3	16	0.85
14 & 15	16.5	16	0.42
Perceptual & cognitive			
16 & 17	24.4	16	0.082
16 & 19	15.9	16	0.46
16 & 20	24.0	16	0.088
17 & 18	15.4	16	0.50
17 & 19	35.3	16	0.004
17 & 20	6.4	16	0.98
18 & 19	14.0	16	0.60
18 & 20	11.4	16	0.79
19 & 20	18.1	16	0.32

evidence of more local dependence than has been included in the GLLRMs.

Tests of no DIF

According to the GLLRMs, there are no problems with DIF in the EFA subscales. STable IV and STable V show the tests of no DIF supporting these claims.

STable IV. Conditional likelihood ratio tests of no differential item functioning (DIF) relative to sex

Items	χ^2	Df	*p*-value
Vegetative			
1	3.6	4	0.47
2	2.8	4	0.59
3	0.6	4	0.96
4	5.9	4	0.21
Facial-oral functions			
5	2.6	4	0.62
6	13.4	4	0.010
7	9.9	4	0.042
8	4.8	4	0.31
Sensorimotor			
9	3.1	4	0.54
10	3.6	4	0.46
11	2.7	4	0.61
12	4.0	4	0.40
14	6.6	4	0.16
15	6.6	4	0.16
Perceptual & cognitive			
16	2.3	4	0.69
17	5.6	4	0.23
18	7.5	4	0.11
19	2.5	4	0.64
20	5.6	4	0.23

Again, there are a few marginally significant test results that are dismissed by the Benjamini & Hochberg (18) procedure. It is worth noting that the global tests of no DIF found weakly significant evidence of DIF relative to Sex for the FOF and SMF subscales (see Table V in the original paper). Since STable IV shows weak evidence of DIF for items 6 and 7 relative to Sex, we recommend that special attention is given to DIF for these items the next time data are collected and analysed on the EFA. For now, we conclude that the evidence of DIF is not strong enough to include in the GLLRM model for the FOF subscale.

Person estimates

STable VI shows Warm's (17) weighted maximum likelihood estimates of the person parameters of the GLLRMs fitting data, together with information about bias and standard errors of measurement (SEMs). Since we assume that many EFA users will prefer to use total raw scores instead of the person parameter estimates, STable VI also includes information on the standard errors of these scores regarded as estimates of the true expected scores measured without error.

Person parameters have values on interval scales with arbitrary origins and units. Since many users are uncomfortable with negative parameters, it is customary to re-scale the parameters by changing the origin and unit so that all parameters are positive. The routine way to do this is so that parameters lie within a 0–1000 or 0–100 interval. Since such transformations have been known to lead to misinterpretation of the re-scaled parameters in terms of percentages, we re-scale the EFA parameters so that the ranges of the re-scaled parameters are the same as the ranges of the summated scores. Since

STable V. Conditional likelihood ratio tests of no differential item functioning (DIF) relative to Age

Items	χ^2	Df	*p*-value
Vegetative			
1	8.6	8	0.37
2	4.0	8	0.77
3	11.0	8	0.20
4	4.3	8	0.83
Facial-oral functions			
5	14.7	8	0.065
6	8.5	8	0.389
7	7.5	8	0.482
8	18.0	8	0.021
Sensorimotor			
9	21.1	8	0.007
10	7.5	8	0.48
11	18.5	8	0.018
12	3.4	8	0.91
14	8.8	8	0.36
15	8.1	8	0.42
Perceptual & cognitive			
16	9.2	8	0.33
17	16.3	8	0.038
18	12.4	8	0.14
19	15.2	8	0.05
20	6.8	8	0.56

Supplementary material to article by I. Poulsen et al. "Validation of the early functional abilities scale: an assessment of four dimensions in early recovery after traumatic brain injury"

STable VI. Weighted maximum likelihood estimates of person parameters

	Vegetative functions				Facial-oral functions				Sensorimotor functions[1]				Perceptual and cognitive Function			
Score	Person param.	SEM	Bias	Score SEM	Person param.	SEM	Bias	Score SEM	Person param.	SEM[2]	Bias	Score SEM	Person param.	SEM	Bias	Score SEM
0	**−5.64**	1.22	0.647		**−4.91**	1.00	0.561		**−3.91**	0.72	0.443		**−7.38**	1.22	0.649	
1	**−3.48**	1.22	0.092	0.71	**−3.30**	1.06	0.080	0.82	**−2.96**	0.71	0.074	1.02	**−5.22**	1.15	0.077	0.74
2	**−2.04**	1.01	−0.002	0.96	**−2.24**	0.95	0.027	1.02	**−2.45**	0.69	0.018	1.36	**−4.04**	0.96	−0.010	1.01
3	**−1.26**	0.84	−0.001	1.20	**−1.40**	0.78	−0.005	1.22	**−2.04**	0.64	0.008	1.55	**−3.35**	0.84	−0.007	1.22
4	**−0.71**	0.71	0.002	1.42	**−0.89**	0.68	−0.015	1.47	**−1.68**	0.59	0.006	1.69	**−2.80**	0.75	−0.001	1.35
5	**−0.29**	0.60	−0.003	1.64	**−0.56**	0.62	−0.009	1.69	**−1.36**	0.55	0.004	1.81	**−2.31**	0.70	0.000	1.44
6	**0.03**	0.52	−0.012	1.87	**−0.28**	0.56	−0.002	1.82	**−1.08**	0.52	0.003	1.93	**−1.86**	0.67	−0.001	1.50
7	**0.26**	0.47	−0.016	2.11	**0.00**	0.52	0.000	1.90	**−0.83**	0.49	0.003	2.05	**−1.45**	0.65	−0.001	1.54
8	**0.42**	0.45	−0.015	2.32	**0.26**	0.50	−0.004	1.97	**−0.60**	0.46	0.002	2.17	**−1.05**	0.65	0.002	1.55
9	**0.56**	0.44	−0.019	2.46	**0.50**	0.49	−0.007	2.04	**−0.40**	0.43	−0.001	2.30	**0.64**	0.66	0.005	1.52
10	**0.69**	0.45	−0.003	2.51	**0.72**	0.50	−0.004	2.07	**−0.22**	0.40	−0.005	2.44	**−0.21**	0.68	0.006	1.47
11	**0.81**	0.47	0.004	2.45	**0.92**	0.52	0.002	2.03	**−0.06**	0.38	−0.009	2.61	**0.27**	0.70	0.003	1.42
12	**0.94**	0.50	0.009	2.25	**1.13**	0.56	0.011	1.90	**0.07**	0.36	−0.010	2.77	**0.80**	0.72	−0.002	1.39
13	**1.10**	0.55	0.007	1.89	**1.38**	0.62	0.018	1.67	**0.17**	0.35	−0.009	2.92	**1.34**	0.72	0.006	1.38
14	**1.31**	0.63	−0.015	1.35	**1.71**	0.72	0.015	1.33	**0.27**	0.35	−0.006	3.02	**1.87**	0.73	0.006	1.37
15	**1.74**	0.84	−0.125	0.68	**2.29**	0.87	−0.037	0.89	**0.36**	0.36	−0.001	3.06	**2.38**	0.75	0.004	1.35
16	**5.01**	1.70	−0.882		**4.02**	1.04	−0.598		**0.44**	0.37	0.004	3.02	**2.92**	0.78	0.000	1.29
17									**0.53**	0.38	0.010	2.90	**3.52**	0.86	0.007	1.18
18									**0.63**	0.41	0.015	2.69	**4.26**	0.98	0.008	0.99
19									**0.74**	0.44	0.019	2.39	**5.48**	1.15	−0.075	0.74
20									**0.89**	0.50	0.021	2.02	**7.56**	1.18	−0.637	
21									**1.13**	0.59	0.015	1.62				
22									**1.54**	0.75	0.007	1.22				
23									**2.38**	0.99	0.080	0.83				
24									**4.30**	1.12	0.626					

[1]Summated score of sensorimotor function ranges from 0 to 24 after elimination of Item 13 "Standing". Scores and estimates are in bold.
SEM: standard error of measurement; param.: parameter.

JRM Journal of Rehabilitation Medicine

person parameters are measured on interval scales, the correspondence between the observed scores and the associated person parameters clearly shows that the summated scores do not provide interval-scaled measures.

STable VII shows the re-scaled person parameters. To remind the reader that the re-scaled parameters are estimates, STable VIII includes information on standard errors and bias of the re-scaled parameters.

STable VII. Re-scaled estimates of person parameters

	Vegetative functions				Facial-oral functions				Sensorimotor functions[1]				Perceptual and cognitive functions			
Score	Person param.	SEM	Bias	Score SEM	Person param.	SEM	Bias	Score SEM	Person param.	SEM[2]	Bias	Score SEM	Person param.	SEM	Bias	Score SEM
0	**0.0**	1.82	0.971		**0.0**	1.80	1.005		**0.0**	2.12	1.295		**0.0**	1.62	0.869	
1	**3.2**	1.83	0.139	0.71	**2.9**	1.90	0.143	0.82	**2.8**	2.07	0.217	1.02	**2.9**	1.54	0.103	0.74
2	**5.4**	1.52	−0.004	0.96	**4.8**	1.70	0.049	1.02	**4.3**	2.02	0.052	1.36	**4.5**	1.28	−0.014	1.01
3	**6.6**	1.27	−0.002	1.20	**6.3**	1.40	−0.009	1.22	**5.5**	1.87	0.025	1.55	**5.4**	1.12	−0.010	1.22
4	**7.4**	1.06	0.003	1.42	**7.2**	1.21	−0.027	1.47	**6.5**	1.72	0.018	1.69	**6.1**	1.00	−0.001	1.35
5	**8.0**	0.90	−0.005	1.64	**7.9**	1.10	−0.016	1.69	**7.4**	1.61	0.012	1.81	**6.8**	0.93	0.000	1.44
6	**8.5**	0.78	−0.18	1.87	**8.3**	1.01	−0.003	1.82	**8.3**	1.51	0.009	1.93	**7.4**	0.89	−0.001	1.50
7	**8.9**	0.71	−0.024	2.11	**8.8**	0.94	−0.001	1.90	**9.0**	1.43	0.008	2.05	**8.0**	0.88	−0.001	1.54
8	**9.1**	0.67	−0.022	2.32	**9.3**	0.90	−0.008	1.97	**9.7**	1.34	0.005	2.17	**8.5**	0.87	0.002	1.55
9	**9.3**	0.66	−0.015	2.46	**9.7**	0.88	−0.012	2.04	**10.3**	1.25	−0.004	2.30	**9.0**	0.89	0.006	1.52
10	**9.5**	0.67	−0.004	2.51	**10.1**	0.89	−0.007	2.07	**10.8**	1.17	−0.015	2.44	**9.6**	0.91	0.008	1.47
11	**9.7**	0.70	0.007	2.45	**10.4**	0.94	0.004	2.03	**11.2**	1.10	−0.025	2.61	**10.2**	0.93	0.004	1.42
12	**9.9**	0.75	0.014	2.25	**10.8**	1.01	0.019	1.90	**11.6**	1.05	−0.029	2.77	**11.0**	0.96	−0.003	1.39
13	**10.1**	0.82	0.011	1.89	**11.3**	1.12	0.032	1.67	**11.9**	1.03	−0.026	2.92	**11.7**	0.97	−0.007	1.38
14	**10.4**	0.95	−0.022	1.35	**11.9**	1.29	0.027	1.33	**12.2**	1.03	−0.017	3.02	**12.4**	0.98	−0.008	1.37
15	**11.1**	1.26	−0.187	0.68	**12.9**	1.56	−0.066	0.89	**12.5**	1.04	−0.004	3.06	**13.1**	1.00	−0.005	1.35
16	**16.0**	2.55	−1.325		**16.0**	1.86	−1.071		**12.7**	1.07	0.012	3.02	**13.8**	1.05	0.000	1.29
17									**13.0**	1.12	0.029	2.90	**14.6**	1.15	0.010	1.18
18									**13.3**	1.19	0.045	2.69	**15.6**	1.31	0.011	0.99
19									**13.6**	1.29	0.057	2.39	**17.2**	1.54	−0.101	0.74
20									**14.0**	1.45	0.060	2.02	**20.0**	1.58	−0.852	
21									**14.7**	1.72	0.043	1.62				
22									**16.0**	2.20	−0.020	1.22				
23									**18.4**	2.90	−0.235	0.83				
24									**24.0**	3.28	−1.829					

[1]Summated score of sensorimotor function ranges from 0 to 24 after elimination of Item 13 "Standing". Scores and estimates are in bold.
SEM: standard error of measurement; param.: parameter.

JRM Journal of Rehabilitation Medicine

Supplementary material to article by I. Poulsen et al. "Validation of the early functional abilities scale: an assessment of four dimensions in early recovery after traumatic brain injury"

STable VIII. Means, standard deviations and standard errors of measurement (SEM) of the Early Functional Abilities (EFA) sub-scores

	Target				Study population				
Subscale	Person parameter	SEM	True score	SEM (TS)	Person parameter[a]	SEM[a]	True score[a]	SEM (TS)	Reliability
VF	0.68	0.40	10.0	2.5	–1.24	0.86	4.5	1.4	0.87
FOF	0.72	0.48	10.0	2.1	–0.78	0.75	5.8	1.4	0.88
SMF	0.36	0.33	15.0	3.1	–0.84	0.59	9.3	1.9	0.93
PCF	–1.15	0.65	7.8	1.5	–1.16	0.87	8.1	1.2	0.95

[a]Values refer to mean values in the current study population. SEM: standard error of measurement; TS: true score. VF: vegetative (autonomic) function ; FOF: facio-oral function; SMF: sensorimotor function; PFC: perceptual & cognitive function.

Since the primary purpose of STable VI and STable VII is to show how summated scores can be transformed into interval-scaled measures, the scores and estimates are written in bold in the tables.

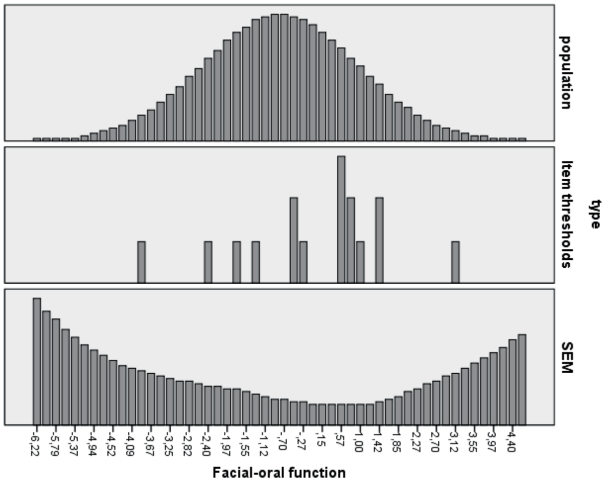
Targeting and item maps

STable VIII provides summary information on the degree to which the EFA subscales target the population of patients, including information on the person parameters (the targets), where measurement error is minimized.

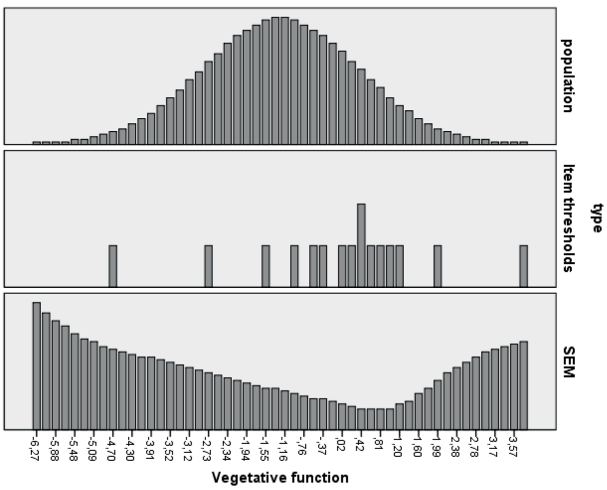
Reliability is high for all for EFA subscales, but the VF, FOF and SMF subscales are somewhat off target, targeting patients with better functioning than the study population (STable VIII). The target of the VF subscale measuring vegetative functions is equal to 0.68, with a target SEM equal to 0.40 and an expected (true) score equal to 10.0. The mean person parameter in the sample is equal to –1.24, corresponding to a true score equal to 4.5, and the mean SEM in the population is equal to 0.86, which is more than twice the target SEM.

The targeting of the PCF subscale is better. The mean person parameter is almost the same as the target and the mean SEM of 0.87 is only 33% larger than the target SEM.

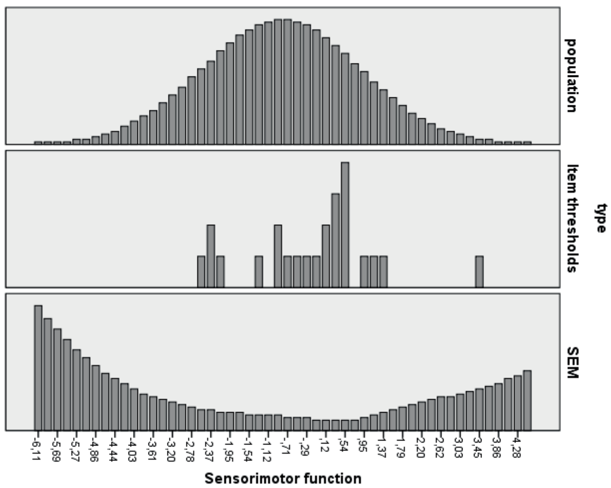
SFigs 1–4 show the item maps comparing the distribution of person parameters with the value of the item thresholds. The third panel of these maps shows how the SEM depends on the person parameters. Except for the PCF subscale, the EFA subscales appear to be somewhat off target, because many patients have parameters below the range of the item thresholds, where measurement error is relatively low. From this point of view, the PCF appears to be the superior subscale. However, in contrast, Table VI shows that the SEM of the PCF subscale is generally larger than the SEM of the other 3



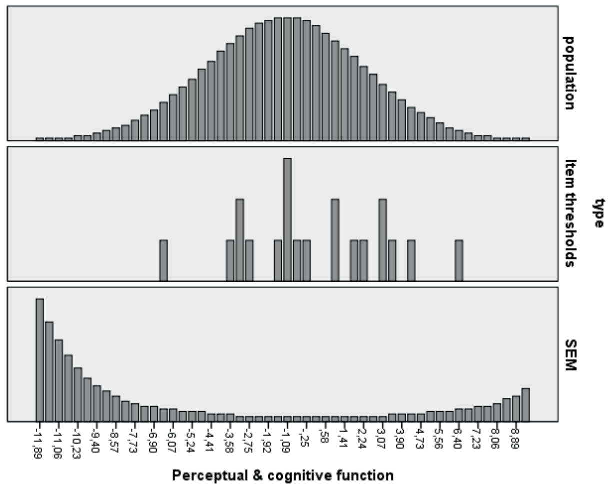
SFig. 2. Distribution of persons and item thresholds together with the standard errors of measurement (SEM) of the facio-oral function (FOF) subscale.



SFig. 1. Distribution of persons and item thresholds together with the standard errors of measurement (SEM) of the vegetative (autonomic) function (VF) subscale.



SFig. 3. Distribution of persons and item thresholds together with the standard errors of measurement (SEM) of the sensorimotor function (SMF) subscale.



SFig. 4. Distribution of persons and item thresholds together with the standard errors of measurement (SEM) of the perceptual & cognitive function and ADL (PCF) subscale.

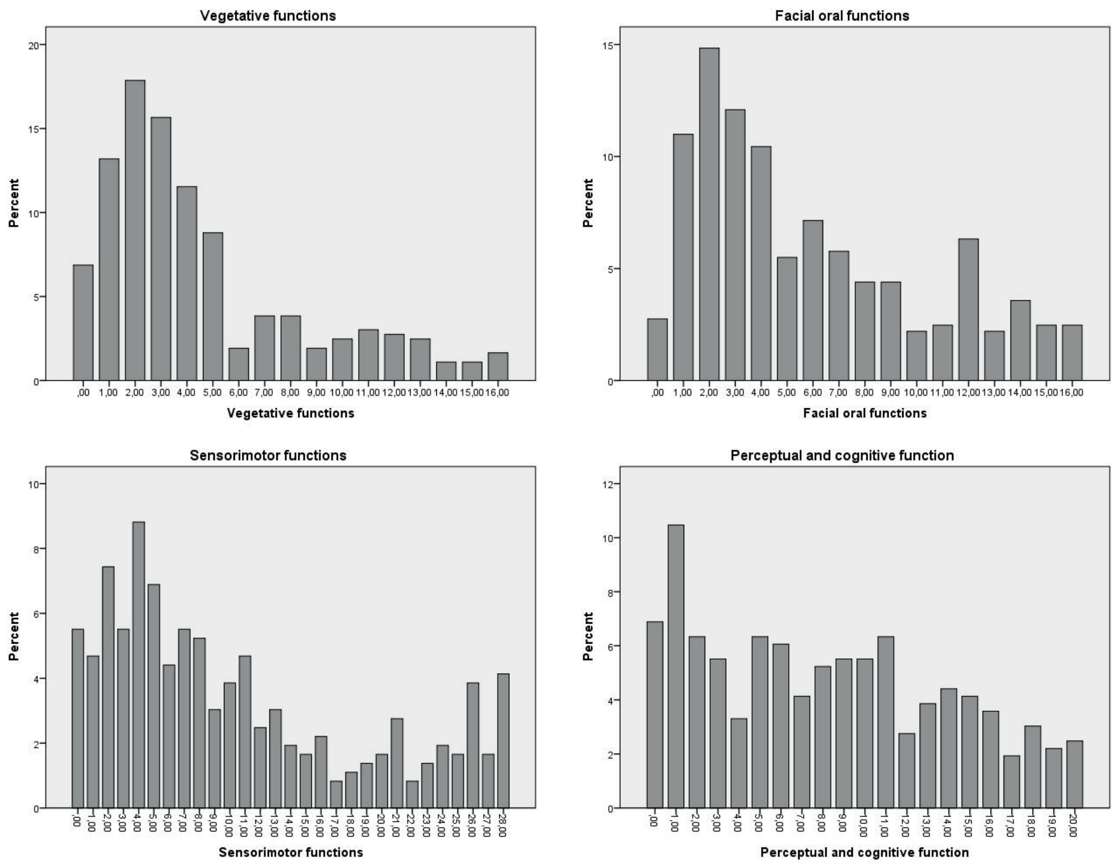
subscales. An explanation of this phenomenon requires careful analysis of the item thresholds of the different EFA items, together with the effects on measurement by local response dependence, which was not part of the current study. For now, we can conclude only that the information on the mean SEMs of the different subscales in Table VII suggests that the better targeting of PCF has been more than offset by the larger SEMs of the PCF.

Distribution of Early Functional Abilities subscores

Finally, SFig. 5 shows the distributions of the EFA subscales. The distributions are skewed, with many patients with low scores illustrating the less than optimal targeting of the EFA. All tests relating EFA subscales to Sex and Age accept that the EFA scores are independent of Sex and Age (all *p*-values are comfortably larger than 0.05).

SUPPLEMENTARY REFERENCES

S1. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielsen & Lydiche; 1960.
S2. Andersen EB. A goodnes of fit test for the Rasch model. Psychometrica 1973; 38: 123–140.
S3. Kreiner S, Christensen K.B. Overall tests of the Rasch. In: Christensen KB, Kreiner S, Mesbah M, editors. Rasch model in health. London: ISTE Ltd & Wiley & Sons; 2013, p. 105–110.
S4. Lord FN, Novick MR. Statistical theory of mental test scores. Reading, MA: Addison-Wesley; 1968, p. 585.
S5. Hamon A, Mesbah M. Questionnaire reliability under the Rasch model. In: Mesbah M, Cole BF, Lee M-LT, editors. Statistical methods for quality of life studies: Design, measurements and analysis. Boston, MA: Springer US; 2002, p. 155–168.



SFig. 5. Distribution of Early Functional Abilities (EFA) subscales.