



ORIGINAL ARTICLE

# The Early Functional Abilities-revised may bridge the gap between the disorder of consciousness and the functional independence scales: evidence from Rasch analysis

Serena CASELLI <sup>1</sup>, Svend KREINER <sup>2</sup>, Aladar B. IANES <sup>3</sup>, Roberto PIPERNO <sup>4</sup>, Fabio LA PORTA <sup>4</sup> \*

<sup>1</sup>Azienda Ospedaliero-Universitaria di Modena, Modena, Italy; <sup>2</sup>Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark; <sup>3</sup>Korian Group, Milan, Italy; <sup>4</sup>IRCSS Istituto Scienze Neurologiche di Bologna, Bologna, Italy

\*Corresponding author: Fabio La Porta, IRCSS Istituto Scienze Neurologiche di Bologna, c/o Ospedale Bellaria, via Altura 6, 40139 Bologna, Italy.  
E-mail: [fabiolaporta@mail.com](mailto:fabiolaporta@mail.com)

*This is an open access article distributed under the terms of the Creative Commons CC BY-NC-ND license which allows users to copy and distribute the manuscript, as long as this is not done for commercial purposes and further does not permit distribution of the manuscript if it is changed or edited in any way, and as long as the user gives appropriate credits to the original author(s) and the source (with a link to the formal publication through the relevant DOI) and provides a link to the license. Full details on the CC BY-NC-ND 4.0 are available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>.*

## ABSTRACT

**BACKGROUND:** There is a tremendous clinical and research need to bridge the gap between disorder of consciousness and functional independence scales with a single unidimensional measure in people with acquired brain injury.

**AIM:** To calibrate an essentially unidimensional subset of items from the Italian Early Functional Abilities (EFA), demonstrating internal construct validity and sufficient reliability for individual patient measurement.

**DESIGN:** Multicenter observational cross-sectional study.

**SETTING:** Inpatients from 11 different Italian Rehabilitation centers.

**POPULATION:** Three hundred sixty-two adult patients with a disorder of consciousness due to an acquired brain injury.

**METHODS:** The Italian version of EFA was administered to the sample and then submitted to Mokken analysis, Confirmatory Factor Analysis, Rasch analysis, Confirmatory Bifactor Analysis, and external construct validity.

**RESULTS:** According to Mokken Analysis (all item scalability coefficients  $H_j$  positive; all item-pair scalability coefficients  $H_{ij} > 0.3$ ; scale coefficient  $H = 0.762$ ), and Confirmatory Factor Analysis (RMSEA=0.081; SRMR=0.048; CFI=0.995; TLI=0.995), the Italian EFA showed a sufficient preliminary unidimensionality. Within Rasch Analysis, a final 12-item solution for the EFA (EFA-R) was calibrated. EFA-R is “essentially unidimensional” according to the following requirements: 1) analysis of residual correlations which supported item essential local independence; 2) a robust correlation between item subsets ( $\rho = 0.950$ ); 3) only 2.1% of cases with significant difference between person parameter estimates by different subscales; 4) an explained common variance equal to 0.916 obtained from a final Confirmatory Bifactor Analysis. It also satisfied invariance requirement (unconditional  $\chi^2_{20} = 9.81$ ;  $P = 0.457$ , conditional class-interval based  $\chi^2_{35} = 33.1$ ;  $P = 0.557$ ), and monotonicity. The reliability (Person Separation Index=0.887) was adequate for person measurements. A practical raw-score-to-measure conversion table based on the EFA-R calibration was devised. Finally, EFA-R strongly correlated with Coma Recovery Scale-Revised ( $\rho = 0.922$ ) and motor FIM™ ( $\rho = 0.808$ ).

**CONCLUSIONS:** EFA-R is an essentially unidimensional subset of 12 items with adequate internal construct validity and sufficient reliability for individual patient measurement under the Rasch Model Theory framework.

**CLINICAL REHABILITATION IMPACT:** EFA-R has the potential to measure people’s functional abilities whose consciousness is improving despite ongoing severe motor-functional impairments during the early stages of rehabilitation. It provides “a measurement bridge” between the disorder of consciousness and the functional independence scales in patients with severe acquired brain injury.

*(Cite this article as: Caselli S, Kreiner S, Ianes AB, Piperno R, La Porta F. The Early Functional Abilities-revised may bridge the gap between the disorder of consciousness and the functional independence scales: evidence from Rasch analysis. Eur J Phys Rehabil Med 2022;58:805-17. DOI: 10.23736/S1973-9087.22.07522-0)*

**KEY WORDS:** Brain injuries; Consciousness disorders; Neurological rehabilitation; Outcome assessment (health care); Psychometrics.

Following severe acquired brain injury, patients may experience a state of Disorder of Consciousness (DOC), either in the form of an Unresponsive Wakefulness State (UWS) or a Minimally Conscious State (MCS). After emergence from MCS (eMCS), patients are often quite severely disabled and completely dependent, although they may show signs of further recovery of sensorimotor and cognitive functions.<sup>1</sup> These early functional changes on emergence from DOC typically occur below the measurement range of functional independence scales (*e.g.*, Barthel Index [BI] and the FIM<sup>TM</sup><sup>1, 2</sup>). Differently, scales for DOC (*e.g.*, Coma Recovery Scale-Revised [CRS-R]<sup>2</sup>) are equally unable to measure these functional changes, as they usually occur above their measurement range.<sup>3</sup> In other words, there is a measurement gap between DOC scales and functional independence scales, as both cannot measure the recovery of early functional changes taking place on emergence from DOC.

To bridge this measurement gap, Heck *et al.* developed the Early Functional Abilities scale (EFA) in 2000.<sup>1, 3</sup> EFA aims to assess people's functional abilities whose consciousness is improving despite ongoing severe motor-functional impairments occurring during the early stages of rehabilitation. The scale describes clinically observable changes concerning purposeful activities, illness and disability awareness, and the ability to comply with the necessary medical, nursing, and therapeutic interventions.<sup>3</sup> The psychometric properties of the EFA scale have been studied in Danish and German populations<sup>4</sup> in terms of concurrent validity with other instruments (CRS-R, FIM<sup>TM</sup>, BI, and others),<sup>3, 5</sup> predictive validity related to rehabilitation outcomes, like morbidity and length of stay,<sup>5</sup> inter-rater reliability,<sup>1</sup> and combined assessment with FIM<sup>TM</sup>.<sup>4, 6</sup> In 2018, Poulsen *et al.*<sup>7</sup> assessed the internal construct validity, reliability, and measurement precision of EFA in patients with traumatic brain injury (TBI) within the Rasch Measurement Theory (RMT) framework. They provided evidence that measurement by the different EFA subscales is valid, objective, and reliable. However, their analyses rejected unidimensionality, and thus they did not recommend summarizing the four subscale measures into a total EFA scale.<sup>7</sup>

Notwithstanding Poulsen *et al.*'s findings, there is an urgent clinical and research need to bridge the gap between DOC scales and functional independence scales with a single unidimensional measure. Addressing such issues, Stout<sup>8, 9</sup> introduced the notion of "essential unidimensionality." This term describes those situations where multidimensional measurement functions as if it were unidimen-

sional for practical purposes. Unfortunately, Poulsen *et al.* did not address this issue.

For this reason, this study aimed to investigate whether it was possible to select a valid content subset of items (Early Functional Abilities-Revised, EFA-R) from the four EFA-dimensions, providing an essentially unidimensional measurement of early functional ability. In doing so, we applied the perspective of the RMT framework to ensure that the revised scale had adequate internal construct validity and sufficient reliability for individual patient measurements.

## Materials and methods

### Setting and patients

Data were collected retrospectively across eleven Italian centers, including eight rehabilitation wards, one intermediate care facility, and two nursing homes (NH), between July 2009 and February 2018. We included in this study all patients aged  $\geq 18$  years with a diagnosis of DOC due to an acquired etiology on admission to these units. Exclusion criteria were: pre-existing neurological degenerative pathologies and/or concurrent illnesses (*e.g.*, tumors) likely to compromise survival within six months. Patients with severe medical instability were also temporarily excluded from enrolment until their clinical conditions had not improved. As all patients had a variable number of EFA assessments during their hospital stay, we randomly selected one evaluation only. Thus, each patient was represented once in the dataset to avoid the risk of time dependency.<sup>10</sup>

Data collection complied with the ethical principles outlined in the Helsinki Declaration.<sup>11</sup> Indeed, upon admission, the legal representatives of the patients had given their informed consent for using anonymously any assessment material generated during the hospital stay also for research purposes. Therefore, as the scale was part of the routine assessment protocol for inpatients, we did not submit the study to the local Ethics Committee.

### Instruments

The EFA Scale<sup>3</sup> includes 20 items grouped in 4 subscales: vegetative functions, facio-oral functions, sensorimotor functions, and perceptual and cognitive functions. Each item is rated on a five-point scale (1="no function," 2="severe disturbance," 3="moderate disturbance," 4="slight disturbance," 5="normal"), providing a total score ranging from 20 to 100.<sup>5, 7</sup> The English version of the EFA scale<sup>5</sup> was translated into Italian and back-translated into English using standardized methods.<sup>12</sup>

The Italian version of the EFA was administered by the rehabilitation team members, including nurses (vegetative functions items), speech and language therapists (facio-oral functions), physiotherapists (sensorimotor functions), and occupational therapists (perceptual and cognitive functions). In addition, all raters were adequately trained based on the initially written scoring guidelines of the EFA scale to minimize inter-rater variability.

Where available, the following other scales collected for each person at the same time as the EFA assessment were used for sample description and external validity purposes:

- CRS-R: a bedside standardized neuro-behavioral assessment tool for individuals with DOC, incorporating and operationalizing all current diagnostic criteria for UWS, MCS, and eMCS.<sup>13</sup> It consists of six items addressing auditory, visual, motor, oromotor/verbal, communication, and arousal functions, yielding a total score from 0 to 23.<sup>2</sup>
- FIM™: one of the most widely used outcome measures to assess independence in activities of daily living in neurological rehabilitation.<sup>14</sup> Only the motor FIM™ (mFIM™ of 13 items) was administered. All items are rated on a 7-point ordinal scale, with higher scores indicating more independence (total score 7-91).

#### Definition of “essential unidimensionality”

• A multidimensional scale is defined as “essentially unidimensional” if the measurement functions as if it only depends on a single latent trait variable for practical purposes. Since multidimensionality is known to induce evidence of local dependence among items, Stout<sup>9, 10</sup> defined “essential unidimensionality” in terms of “essential local independence.” He required that the average of the partial correlations between items, conditionally given the latent trait, must be a small number that disappears as the number of items increases towards infinity. In our paper, we adopted this definition, insisting that the average of the so-called “residual correlations” had to be a small number, *e.g.*, less than 0.10 or less than 0.20. We referred to Marais<sup>15</sup> and Christensen *et al.*<sup>16</sup> for information on the response residuals calculated during Rasch analyses. Since response correlations are known to be biased with expected values less than zero under the Rasch model, they suggested using adjusted correlations, defined by the observed correlations minus the average of the observed correlations.<sup>15, 16</sup> Hence, during our assessment of “essential unidimensionality,” we required that the average of the absolute values of the adjusted correlations was less than 0.20.

• In addition to Stout’s requirement, we added three different requirements to “essential unidimensionality”:

• First, correlations between scores measuring essential unidimensional measurement had to be high. Test-retest reliability calculation assumes that repeated measurements are conditionally independent given the latent trait. Essential unidimensionality assumes that subscales should measure close to the same trait and should be conditionally independent given the latent trait. For this reason, the correlation between the subscales plays the same role as the test-retest correlation. Since a test-retest correlation equal to 0.9 is regarded as excellent reliability in supporting the clinical application of test results, we insisted that the correlation between two subscales ( $r$ ) must be close to 90%.<sup>17</sup>

• Second, a Confirmatory Bifactor Analysis (CBA), defined by a general factor and several specific factors, should show that the specific factors have little effect on test scores above and beyond the effect of the general factor. One commonly used requirement for “essentially unidimensionality” is that the general factor explains close to 90% of the variance of the test scores (explained common variance, ECV).<sup>18, 19</sup>

• The third requirement considers “essential unidimensionality” from a different point of view. One way to test unidimensionality in a Rasch model is to compare estimates of person parameters defined by different subscores.<sup>20, 21</sup> Only if the person estimates are significantly different in close to 5% of the cases do we conclude that unidimensionality is satisfied.<sup>20, 21</sup> In large sample studies, the difference between the expected 5% and the observed number of cases needs not be much larger than 5% to provide significant evidence against unidimensionality because of the power of the statistical tests. In such cases, we claimed that measurement was “essentially unidimensional” if the number of cases with significantly different estimates of person parameters was less than 10%.

#### Psychometric analyses

The following analyses were undertaken:

- Sample descriptive statistics (persons and EFA items);
- Initial analysis of the scalability of EFA items by Mokken analysis (MA);
- Preliminary assessment of local dependence and dimensionality of EFA items using Confirmatory Factor Analysis (CFA);
- Item analysis of EFA by polytomous Rasch model;
- Evaluation of the ECV of the final EFA subset by Confirmatory Bifactor Analysis (CBA);
- Concurrent validity (external construct validity) of the final EFA subset.

### Descriptive statistics of persons and items

Descriptive statistics for persons' main demographic and clinical variables (age, gender, etiology, number of days since lesion, setting) and analyses of the frequency of score categories, missing data, and item-to-total correlations (Spearman rho) were performed.

### MA

To obtain initial information on the scalability of EFA items, we performed a MA of the EFA, which is a scaling procedure for ordinal items, based on the Monotone Homogeneity Model.<sup>22</sup> It assumes the unidimensionality of the latent trait, the monotonicity, and the local independence of responses. It can be used to partition a set of items into Mokken scales using an automated item selection procedure.<sup>22</sup> We used the following indicators:

- item scalability coefficient  $H_j$  (normed covariance between the item and the rest scores<sup>22</sup>): values should be  $\geq 0.3$  (recommended default value of positive lower bound  $c$ );
- item-pair scalability coefficients  $H_{ij}$  (normed covariance between the item scores): values should be positive for items belonging to the same Mokken scale;<sup>23</sup>
- scalability coefficient  $H$ : indicates the overall quality of a scale (*i.e.*, the degree to which the test data follow a perfect Guttman pattern<sup>23</sup>).

At the end of the procedure, the analysis shows the number of scales needed for scaling all items. Should the automated algorithm estimate the need for more than one scale to accommodate all the items, we would consider this information on the item scalability in the following analysis steps.

### Confirmatory factor analysis

To assess the fit of the EFA items to a unidimensional factor analysis model, we calculated several fit statistics and tests of local dependence<sup>24, 25</sup> referred to as modification indices (MI).<sup>26, 27</sup> Since items are ordinal categorical variables, we used polychoric correlations<sup>27, 28</sup> during the analysis. The indicators of fits included:

- the Root Mean Square Error of Approximation (RMSEA): values  $\leq 0.08$  indicative of "mediocre fit" but sufficient for a preliminary assessment of dimensionality;<sup>29</sup>
- the Standardized Root Mean square Residual (SRMR): values  $\leq 0.08$  indicative of "adequate fit"<sup>30</sup>;
- the comparative fit index (CFI) and the non-normed fit index (Tucker-Lewis Index – TLI): values  $> 0.95$  [0, 1] were considered acceptable.<sup>26</sup>

Evidence against a CFA model does not necessarily im-

ply evidence against a polytomous Rasch model. A CFA model of polychoric correlations approximates a so-called "graded response model"<sup>31</sup> and imposes restrictions on the distributions of the latent variable that differ from the assumptions of the Rasch model. However, evidence of the CFA indicators model and local dependence can be regarded as evidence against unidimensionality that probably extends to a Rasch model.

### Rasch analysis

Following the above analyses, all 20 EFA items were fitted to the Rasch model.<sup>32</sup> We use the partial credit parameterization of the model and address the following measurement issues:<sup>33, 34</sup>

- Internal construct validity, which included: 1) the assessment of item homogeneity or invariance;<sup>20, 32, 35</sup> 2) the adherence to a probabilistic Guttman pattern;<sup>20, 35</sup> 3) monotonicity;<sup>20, 36</sup> 4) local independence;<sup>15, 16, 36, 37</sup> 5) unidimensionality;<sup>20, 21, 36</sup> and 6) absence of differential item functioning (DIF) relative to subgroups defined by age, gender, etiology, days since lesion, setting.<sup>20, 36</sup>
- Targeting,<sup>25, 38</sup> expressed as floor and ceiling effects<sup>39</sup> and targeting index.<sup>39</sup>
- Assessment of reliability, represented by the Person Separation Index (PSI) and the Cronbach's  $\alpha$ .<sup>20, 35, 40</sup>

Where these assumptions failed, an iterative phase involving item modifications (item rescaling, item grouping, item splitting, item deleting) aimed at finding a solution that satisfied the model requirements was undertaken.<sup>29, 32</sup>

In the case of a final two-testlet solution, conditional total item-trait interaction chi-squares were calculated because the unconditional ones are known to be unreliable for sample sizes of 200 or more. Compared to this, the conditional fit statistics remain reliable for sample sizes  $\leq 2,000$ .<sup>41</sup>

Should DIF be detected, the influence of the item/testlet splitting on the person estimates would be tested using the procedure presented by Maritz *et al.*<sup>42</sup> After item/testlet splitting, we anchor the "splitted" solution on the "un-splitted" one, using an item/testlet free from the DIF, and compare the person estimates of the two solutions, calculating an effect size (Cohen's  $d$ ) of the paired t-test of the difference. Should Cohen's  $d$  be  $< 0.2$ , it would be considered negligible; thus, the DIF would not be adjusted for.<sup>42</sup> Otherwise, the "splitted" solution would be the final one.<sup>42</sup>

Should a final fitting solution following the above modifications be found, its total score would be transformed into interval-level measurements, whose unit is the logit.<sup>20, 32, 35</sup>



### Confirmatory bifactor analysis

Finally, because one of the requirements of “essential unidimensionality” is that the general factor explains close to 90% of the variance of the test scores (ECV),<sup>18, 19</sup> we performed a CBA of the final EFA subset (EFA-R).

### Concurrent validity (external construct validity)

We calculated pairwise Spearman’s correlations ( $\rho$ ) between EFA-R estimates and CRS-R and mFIM™ total scores to test the hypothesis that the EFA-R could bridge the gap between DOC and functional scales. Indeed, we hypothesized that the degree of correlation of EFA-R estimates with both the CRS-R and mFIM™ scores would be similar and higher than the direct correlation between CRS-R and mFIM™.

### Statistical notes, softwares, and sample size issues

All descriptive statistics and correlations were performed using SPSS software (SPSS. Version 13 for Windows; www.spss.com). We undertook CFA and CBA for ordinal data using the Mplus software (Mplus version 6.0. Muthen & Muthen, 1998–2010; www.statmodel.com). MA was performed using R (R version 3.5.0 [2018-04-23]). We carried out the Rasch analysis using the RUMM2030 software (version 5.52 for Windows. RUMM Laboratory Pty Ltd, Perth, Australia; 1997–2016; www.rummlab.com), estimating that a sample size of 362 cases would be sufficient to estimate item difficulty to  $\pm 0.5$  logits, with a confidence of 99%, irrespective of the targeting of persons to the items.<sup>43</sup> A significance value of 0.05 was used throughout and corrected for the number of tests by Bonferroni correction. For all reliability analyses, cutoffs of  $>0.70$  and  $>0.90$  were considered adequate for group and individual person measurements, respectively.<sup>24</sup> An *ad hoc* Excel 2007™ application “RUMM logbook” (La Porta F, on behalf of the European Rasch Research & Teaching Group. RUMM Logbook v1.9.5. Bologna, Italy; 2018) was developed using Microsoft Visual Basic™ macros (Microsoft Excel for Windows, version 12.0)<sup>33</sup> to facilitate the interpretation of each Rasch analysis results. A free copy of this application can be requested from the corresponding author (fabiolaporta@mail.com). Finally, to facilitate the interpretation of the absolute values of correlation and effect size coefficients, a modified version of the cutoff criteria provided by Pallant<sup>44</sup> was adopted: negligible: 0–0.09; weak: 0.10–0.29; moderate: 0.30–0.49; strong: 0.50–0.79; very strong:  $\geq 0.80$ .

### Data availability statement

The raw data supporting the conclusions of this article are available for download at Zenodo.org (according to the license Creative Commons Attribution 4.0 International) from the following link: <https://doi.org/10.5281/zenodo.7178651>.

## Results

### Psychometric analyses

#### Descriptive statistics

All observations were collected on a convenience sample of 362 patients. 268 patients (74%) were enrolled within the Rehabilitation Units, 49 (13.5%) in two Nursing Homes, and the remaining 45 patients (12.5%) in Severe Brain Injury Units. The main demographic, clinical, and scale descriptive statistics are summarized in Table I. The median EFA total score for the whole observation sample was 37 (range: 20–100; mean=46; SD=22.4). No missing data were present.

The item-to-total correlations were high (median value: 0.851), ranging from 0.385 (EFA01) to 0.905 (EFA07).

#### MA

The automated item selection procedure within the MA showed the scalability of all the 20 items on one single scale. Indeed, all item-pair scalability coefficients  $H_{ij}$ s were positive, and all the item scale coefficient  $H_{js}$  were higher than 0.3. Furthermore, the scalability coefficient for the entire scale  $H$  was equal to 0.762, which qualifies as a “strong scale.”

### Confirmatory Factor Analysis

The baseline Confirmatory Factor Analysis (CFA), undertaken on the whole sample ( $N=362$ ), failed to support the unidimensionality of the scale (RMSEA=0.109; SRMR=0.059; CFI=0.991; TLI=0.990). However, six pairs of items showed large modification indices (EFA01-02, EFA05-06, EFA06-07, EFA09-10, EFA12-15, EFA17-18). After allowing correlation of the errors within the dependent pairs, it was possible to fit a final model almost reaching the intended cutoff for sufficient unidimensionality before Rasch analysis (RMSEA=0.081; SRMR=0.048; CFI=0.995; TLI=0.995).

### Rasch analysis

The base Rasch analysis showed that the scale did not fit the Rasch model (Table II, “Base analysis”), failing the

TABLE I.—*Sample descriptive statistics (N.=362).*

	N	%	Median	Range [min, max]	Mean	SD
<b>Setting</b>						
Rehabilitation Unit	268	74.0				
Nursing Home	49	13.5				
Severe Brain Injury Unit	45	12.5				
Missing values	0	-				
<b>Gender</b>						
Males	236	65.2				
Females	125	34.5				
Missing values	1	0.3				
Age	357	98.6	59.0	[15, 90]	56.2	16.8
Missing values	5	1.4				
<b>Etiology</b>						
Hemorrhagic stroke	157	43.4				
Traumatic brain injury	123	34.0				
Anoxic brain injury	31	8.6				
Ischemic stroke	28	7.7				
Other etiologies*	23	6.3				
Missing values	0	-				
<b>Time since lesion (days)</b>						
Whole sample	344	95.0	42.0	[1, 4341]	205.9	464.9
Rehabilitation	258	75.1	34.0	[1, 479]	53.0	63.5
Nursing home	49	14.2	540.5	[63, 4341]	802.8	790.8
Severe Brain Injury Unit	37	10.7	129.0	[26, 2023]	465.7	673.9
Missing values	18	5.0				
<b>Scales</b>						
EFA	362	100	21	[20, 100]		
CRS-R	175	48.3	10.0	[0, 23]		
mFIM™	217	59.9	13.0	[13, 91]		

\*e.g., meningoenephalitis, poisoning, etc.

SD: standard deviation; EFA: Early Functional Abilities; CRS-R: Coma Recovery Scale-Revised; mFIM™: motor FIM™.

item invariance requirement ( $\chi^2_{80}=628.4$ ;  $P=0.000$ ).<sup>20</sup> Furthermore, two items (EFA01 and EFA02) had fit residuals  $>2.5$ , highlighting model underfit. Beyond this, there were highly significant chi-squares for seven items (violation of the stochastic invariance of the item hierarchy). Also, the scale failed the unidimensionality requirement, as the Proportion of Significant T-test (PST) was 24.7%, and the Lower Bound of Binomial Confidence Interval for proportions (LBBCI) was 23%. Besides, there were disordered thresholds for eleven items (violation of the monotonicity requirement). Finally, twenty-five pairs of items had adjusted residual correlations above the  $Q3$ ,\* critical value proposed by Christensen *et al.*<sup>20</sup> (here set at 0.18 at 95<sup>th</sup> percentile for 20 polytomous items, sample size of 350, latent mean 0), indicating a violation of the local independence requirement.

In the next steps of analysis, we performed several item modifications to achieve a final fitting solution for the scale:

- Eight items (EFA 01,02,03,08,09,11,13,15) were deleted to account for local dependence and misfit to the model (Supplementary Digital Material 1, Supplementary Table I);
- Five items were rescored (EFA 06,10,12,14,16), as they showed disordered thresholds;
- “Testlets” (or super-items) were created between clusters of items that demonstrated some left-over local dependence, obtaining a two-testlet solution (EFA 04-05-06-07-10-12-20: “facio-oral functions and ADL” cluster; EFA14-16-17-18-19: “sensorimotor and cognitive functions” cluster).
- The sensorimotor and cognitive functions testlet was

TABLE II.—*Rasch analysis details for the Early Functional Abilities.*<sup>20</sup>

Analysis description	Fitness to the Rasch Model				Internal Construct Validity Requirements						Reliability and targeting					
	Analysis name	N	$\chi^2_{df}$	p	Cutoff <sup>a</sup>	Unidimensionality		Other ICV requirements				Separation reliability		Targeting		
PST (%) <sup>b</sup>						Lower BCI (%) <sup>b</sup>	T-DT <sup>c</sup>	LD <sup>d</sup>	Q3, <sup>*</sup>	T-DIF <sup>e</sup>	PSI <sup>f</sup>	$\alpha$	PL mean	PL SD	SEM <sup>g</sup>	Targeting index <sup>h</sup>
1 Base analysis	362	628.4 <sub>80</sub>	0.000	0.002	24.7	23	55%	25	0.18	269	0.951	-	-1.134	1.837	0.407	-2.78
2 After deleting	362	94.5 <sub>60</sub>	0.003	0.004	11.3	9.9	41.7%	12	0.15	107	0.940	0.962	-0.355	2.366	0.580	-0.63
3 After rescoring	362	85.6 <sub>60</sub>	0.016	0.004	11.3	9.9	0%	10	0.15	103	0.943	0.960	-0.378	2.539	0.607	-0.64
4 After subtesting	362	9.8 <sub>10</sub>	0.457	0.025	2.1	0.6	0%	0	-	22	0.887	0.911	-0.189	1.386	0.467	-0.40
Recommended values →	-	-	-	-	<5.0 <sup>g</sup>	Lower BCI<5.0	0%	0	-	0	≥0.85 <sup>h</sup>	0	0.000	-	-	[-2, 2]

ICV: internal construct validity; N: sample size;  $\chi^2_{df}$ : unconditional chi-square for model fit and its degrees of freedom; p: Bonferroni-corrected  $\chi^2$  probability value; PST: proportion of significant t-test carried out on the estimates that, within a principal component analysis of residuals, loaded positively and negatively (factor loading  $\geq \pm 0.3$ ) on the first component; BCI: binomial confidence interval for PST; T-DT: percentage of items with disordered thresholds; LD: number of item pairs with adjusted residual correlations above the respective Q3,<sup>\*</sup>; Q3,<sup>\*</sup>: critical value at 95<sup>th</sup> percentile to interpret adjusted residual correlations; T-DIF: total DIF load; PSI: Person Separation Index;  $\alpha$ : Cronbach's alpha; PL mean: person location mean; PL SD: person location standard deviation; SEM: Standard Error of Measurement of the person locations.

Values are mean (SD) or as otherwise indicated.

<sup>a</sup>Bonferroni-corrected P-value, which varies by analysis, and that is used to interpret the corresponding  $\chi^2$  P-value; <sup>b</sup>unidimensionality is considered achieved either when PST is <5% or when the lower bound of its BCI is <5%; <sup>c</sup>the T-DT statistic is calculated as the percentage of items with disordered thresholds out of the total number of items. The values range from zero to 100%, where zero indicates the absence of items with disordered thresholds; <sup>d</sup>LD value indicates the number of item pairs with adjusted residual correlations above the respective Q3,<sup>\*</sup> critical value at 95<sup>th</sup> percentile.<sup>20</sup> For each analysis cycle it was set according to the type and the number of items, the sample size, and the latent mean; <sup>e</sup>the T-DIF summary statistic is calculated as the absolute value of the base ten logarithms of all p-values for uniform and non-uniform DIF across all items and across all person factors, which are below the Bonferroni-corrected p-value. The values range from zero to infinite, where zero indicates no DIF; <sup>f</sup>a value of  $\geq 0.850$  suggests a precision of measurement at the individual level, whereas a value between 0.700 and 0.849 indicates precision only at the group level; <sup>g</sup>SEM is calculated with the formula:  $SD \times (1 - reliability)^{1/2}$ , where SD is person location standard deviation, and reliability is the PSI with extremes; <sup>h</sup>the targeting index is calculated as the ratio between the average person measures and the SEM. Targeting is good and respectively fair, when the average person measure is beyond [-1 +1] and [-2, +2] SEM the average item measure (set by default at 0 logits).

splitted to resolve a uniform DIF for setting (subjects admitted to the severe brain injury units found this super-item systematically more challenging to pass than the others). However, the difference between the paired person estimates of the splitted vs. the un-splitted solution yielded Cohen's d of 0.140. Given this effect size <0.2, the effect of DIF on person estimates was considered negligible, and DIF was not accounted for.<sup>42</sup>

After these modifications, the final 12-item solution for the EFA (EFA-R - Supplementary Digital Material 2, Supplementary Text File 2, Italian version available on re-

quest) showed a satisfying fit to the Rasch model (Table II)<sup>15, 16</sup> The scale was “essentially unidimensional” according to the satisfaction of the above requirements: local independence between item residuals (average of the absolute values of the adjusted correlations close to negligible<sup>44</sup> 0.127, see Table III for all adjusted correlations), a correlation between subtests “r” equal to 0.950, a significant difference between estimates of person parameters “overall PST” equal to 2.1% (LBBCI 0.6%). It also satisfied invariance requirement (unconditional  $\chi^2_{20}=9.81$ ; P=0.457, conditional class-interval based  $\chi^2_{35}=33.1$ ; P=0.557), and

TABLE III.—*Adjusted residual correlations between EFA-R items (N.=362, analysis no. 3).*<sup>15, 16</sup>

	EFA04	EFA05	EFA06	EFA07	EFA10	EFA12	EFA14	EFA16	EFA17	EFA18	EFA19
EFA05	-0.033										
EFA06	0.021	0.277									
EFA07	0.029	0.242	0.543								
EFA10	0.112	-0.066	-0.045	-0.038							
EFA12	0.352	-0.134	0.058	0.011	0.209						
EFA14	-0.099	-0.138	-0.119	-0.144	0.133	0.013					
EFA16	-0.120	-0.041	-0.168	-0.175	-0.014	-0.025	0.286				
EFA17	-0.104	0.110	-0.061	-0.069	-0.067	-0.157	-0.125	0.017			
EFA18	-0.146	0.030	-0.193	-0.125	-0.091	-0.256	0.052	0.220	0.286		
EFA19	-0.203	-0.024	-0.162	-0.110	-0.142	-0.192	0.055	0.115	0.133	0.258	
EFA20	0.099	-0.134	-0.085	-0.124	0.132	0.300	0.087	-0.003	-0.093	-0.164	-0.008

The adjusted residual correlations (defined by the observed correlations minus the average of the observed correlations<sup>15, 16</sup>) between EFA-R items (N.=362, analysis no. 3) were reported in the table.

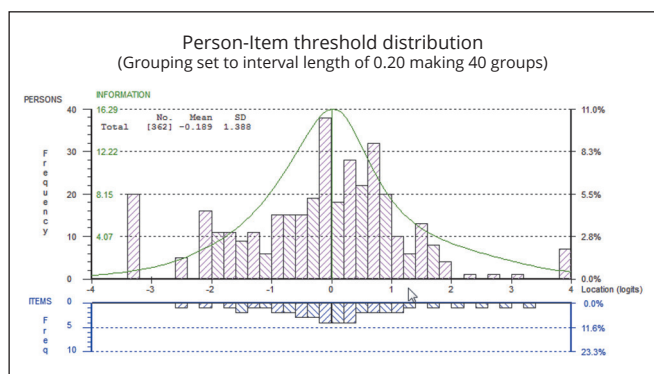


Figure 1.—Targeting of the EFA-R (N.=362). Freq: frequency. In the figure, persons and items are displayed, separated by the logit scale in the upper and lower part of the graph. Grouping set to interval length of 0.20, making 40 groups.

monotonicity (no disordered thresholds). Furthermore, all the subjects’ responses fitted the model. The targeting graph of the EFA-R (Figure 1) highlighted that subjects were spread across eight logits, with negligible floor (5%) and ceiling effects (2%). The mean person ability of -0.189 logits and a targeting index of -0.405 indicated, on average, a proper matching between person ability and item difficulty (set by default at 0 logits).

The separation reliability expressed as PSI and Cronbach’s Alpha was 0.887 and 0.911, respectively, indicating the precision of measurement at the individual level.

The item hierarchy, the scoring model, and the item thresholds are reported in Table IV. Based on the item

calibration, it was possible to construct a table to convert raw scores into interval-level estimates of early functional abilities (Table V).

**Confirmatory Bifactor analysis**

The Confirmatory Bifactor analysis performed on the EFA-R showed an ECV equal to 0.916, satisfying the remaining requirement for “essential unidimensionality.”

**Concurrent validity (external construct validity)**

The EFA-R person estimates correlated “very strongly” both with CRS-R ( $\rho=0.922$ ,  $P<0.000$ ,  $N.=175$ ) and with mFIM™ ( $\rho=0.808$ ,  $P<0.000$ ,  $N.=217$ ) total scores. As hypothesized, the correlation between CRS-R and mFIM™ total scores was “strong” ( $\rho=0.619$ ,  $P=0.000$ ,  $n=134$ ) but lower than the previous ones. Figure 2 shows that the EFA-R scores overlap with the CRS-R ceiling and the mFIM™ floor, overcoming the reciprocal overlap of the two scales.

**Discussion**

Within this study, we calibrated an essentially unidimensional subset of 12 items (EFA-R) from the Italian version of the EFA under the RMT framework. The analysis provided evidence of adequate internal construct validity and sufficient reliability for individual patient measurements. Furthermore, given the strong correlation with CRS-R and mFIM™, EFA-R may successfully bridge the

TABLE IV.—Item parameters, fit statistics, scoring model, and item thresholds for the EFA-R (N.=362, analysis no. 3).

EFA-R items	Subdomain	Item parameters and fit statistics					Scoring model				Item thresholds and standard errors								
		Loc	SE	FR	$\chi^2$	P*	0	1	2	3	4	Thr1	SE	Thr2	SE	Thr3	SE	Thr4	SE
EFA16 – Tactile information	P&C functions	-1.670	0.093	-0.545	6.5	0.273	1	2	2	3	4	-2.451	0.152	0.983	0.145	1.468	0.160		
EFA14 – Voluntary movements	Sensorimotor functions	-1.486	0.091	-0.161	1.9	0.853	1	2	2	3	4	-1.263	0.164	0.298	0.150	0.965	0.162		
EFA18 – Auditory information	P&C functions	-1.325	0.082	-0.817	3.4	0.636	1	2	3	4	5	-2.355	0.185	-0.857	0.153	0.406	0.129	2.807	0.162
EFA10 – Head control	Sensorimotor functions	-1.057	0.089	0.182	4.3	0.501	1	2	3	3	4	-0.822	0.177	-0.536	0.135	1.358	0.155		
EFA17 – Visual information	P&C functions	-0.529	0.081	-0.056	2.2	0.825	1	2	3	4	5	-2.371	0.170	-1.336	0.134	0.918	0.137	2.788	0.190
EFA05 – Oral stimulation and hygiene	Oro-facial functions	-0.491	0.076	-2.057	7.5	0.183	1	2	3	4	5	-2.227	0.158	-0.880	0.137	1.426	0.157	1.682	0.178
EFA06 – Swallowing function	Oro-facial functions	-0.383	0.087	-0.350	13.9	0.016	1	2	3	3	4	-1.413	0.146	-0.063	0.135	1.476	0.163		
EFA19 – Communication	P&C functions	-0.152	0.073	0.835	11.6	0.040	1	2	3	4	5	-1.346	0.163	-0.441	0.160	0.570	0.156	1.217	0.174
EFA07 – Tongue movements and chewing	Oro-facial functions	-0.053	0.071	-1.798	8.2	0.143	1	2	3	4	5	-1.149	0.162	0.027	0.186	0.074	0.156	1.048	0.173
EFA20 – Problem-solving in ADL	P&C functions	2.004	0.081	-1.280	7.2	0.204	1	2	3	4	5	-1.811	0.149	-1.136	0.161	-0.285	0.187	3.232	0.545
EFA12 – Transfers	Sensorimotor functions	2.499	0.103	-0.528	2.6	0.755	1	2	2	3	4	-2.166	0.131	-0.350	0.182	2.516	0.521		
EFA04 – Excretory functions	Autonomic functions	2.644	0.091	0.837	16.1	0.006	1	2	3	4	5	-1.261	0.152	-0.542	0.200	0.848	0.333	0.955	0.485

SE: standard error; EFA-R: Early Functional Abilities-Rasch; P:  $\chi^2$  probability; Thr: threshold; ADL: activities of daily living; P&C: perceptual and cognitive. EFA-R items are ordered by progressively increasing the difficulty from top to bottom. Locations and thresholds are expressed in logits. The degrees of freedom for each  $\chi^2$  were 5 for all items.

\*The Bonferroni-corrected P value indicating statistical significance at the 0.05 level was 0.004.



TABLE V.—Raw-score-to-measure-estimates conversion table for the EFA-R.

Raw score	Logit scale	±95% CI	0-100 scale	±95% CI
0	-3.217	2.097	0.0	14.8
1	-2.588	1.488	8.9	10.5
2	-2.154	1.172	15.0	8.3
3	-1.854	1.011	19.3	7.2
4	-1.621	0.904	22.6	6.4
5	-1.429	0.825	25.3	5.8
6	-1.265	0.764	27.6	5.4
7	-1.124	0.711	29.6	5.0
8	-0.999	0.674	31.4	4.8
9	-0.887	0.641	33.0	4.5
10	-0.786	0.613	34.4	4.3
11	-0.693	0.590	35.7	4.2
12	-0.607	0.568	36.9	4.0
13	-0.527	0.551	38.1	3.9
14	-0.452	0.537	39.1	3.8
15	-0.381	0.523	40.1	3.7
16	-0.314	0.514	41.1	3.6
17	-0.249	0.504	42.0	3.6
18	-0.187	0.496	42.9	3.5
19	-0.127	0.490	43.7	3.5
20	-0.068	0.488	44.5	3.5
21	-0.011	0.486	45.4	3.4
22	0.047	0.486	46.2	3.4
23	0.105	0.488	47.0	3.5
24	0.163	0.490	47.8	3.5
25	0.221	0.498	48.6	3.5
26	0.282	0.506	49.5	3.6
27	0.346	0.515	50.4	3.6
28	0.413	0.529	51.4	3.7
29	0.485	0.547	52.4	3.9
30	0.563	0.566	53.5	4.0
31	0.648	0.592	54.7	4.2
32	0.744	0.621	56.0	4.4
33	0.851	0.659	57.5	4.7
34	0.973	0.702	59.3	5.0
35	1.114	0.753	61.3	5.3
36	1.279	0.811	63.6	5.7
37	1.474	0.882	66.4	6.2
38	1.705	0.960	69.6	6.8
39	1.978	1.053	73.5	7.4
40	2.298	1.164	78.0	8.2
41	2.679	1.327	83.4	9.4
42	3.184	1.629	90.6	11.5
43	3.852	2.199	100.0	15.6

95%CI: 95% confidence interval (equal to 1.96 standard errors of measurement). Person estimates are expressed in logits and 0-to-100 scale.

measurement gap between DOC and functional independence scales, in agreement with other published results on samples of people with severe acquired brain injury.<sup>3, 4, 6, 7</sup>

The 362-person sample, enrolled across 11 different Italian rehabilitation and chronic care centers, included patients within the full spectrum of DOC (from UWS to emergence from MCS) and fully emerged from it. Furthermore, unlike the previous works,<sup>7</sup> our sample included not only persons with TBI but also with hemorrhagic and ischemic stroke and anoxic brain injury. Finally, it considered persons at different rehabilitative stages, from patients who had just left ICU to those already discharged to nursing homes. Furthermore, at least 50% of the patients assessed in a Rehabilitation unit entered this setting, on average, 53 days from the event, highlighting a longer length of stay in the acute phase due to the severity and complexity of their clinical conditions. Given these considerations, the sample can represent the population recovering early functional abilities following a severe brain injury.

The item classical descriptive statistics, the initial information on item scalability with MA, and the preliminary assessment of the unidimensionality of the EFA using CFA provided complementary evidence. According to MA, all the items represented a single robust unidimensional scale solution. Differently, it was possible to achieve a fitting unidimensional solution within the CFA only after allowing the correlation of errors between several locally dependent item pairs. In addition, the item-to-total correlations analysis identified EFA01 (Vegetative stability) as the item that contributes less to the operational definition of the construct “early functional abilities.” Clinically, it may be explained because this construct, as defined in item EFA01, is related to the stability and supervision of the patient at rest and when stimulated. Considering the varied composition of the sample, it represents a fundamental prerequisite for developing early functional abilities. It seems a different construct from the latter, and it

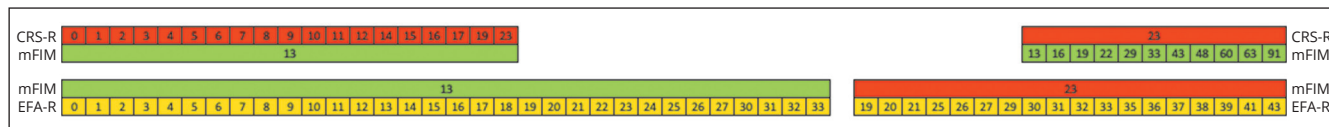


Figure 2.—EFA-R “bridges the measurement gap” between CRS-R and mFIM™. CRS-R: Coma Recovery Scale-Revised; mFIM™: motor FIM™; EFA-R: Early Functional Abilities-Revised. As shown in the left part of the figure, there are 18 CRS-R score levels corresponding to the mFIM floor score compared to the 30 EFA-R score levels. Similarly, there are more EFA-R score levels (18) at the ceiling of the CRS-S in comparison to the 11 mFIM™ score levels (right part of the figure). Thus, although CRS-R and mFIM™ have a partial overlap, EFA-R allows a better separation of subjects both at the ceiling and at the floor of mFIM™ and CRS-R, respectively. Therefore, when a patient obtains an extreme score near the ceiling of the CRS-R, it may be more appropriate to administer the EFA-R scale, as it better discriminates the subject’s ability given its greater score progression and granularity.

has already been recovered at discharge from ICU, where patients have started their rehabilitative treatment.

We performed the internal construct validity analysis on the full item set in the Rasch analysis context, assuming that a subset of items across the four subscales could be essentially unidimensional to measure a single underlying construct. However, it highlighted several violations of the specific requirements of the Rasch model. To account for these violations, we followed a different analytical strategy from Poulsen *et al.*<sup>7</sup> Indeed, as their analysis was based on each subscale, their approach was more conservative, as only one item (EFA13 – sensorimotor subscale) was deleted. On the other hand, we removed several items (EFA 01, 02, 03, 08, 09, 11, 13, 15), given their serious misfit to the model and multidimensionality. It is worth noting that although this action removed 28 thresholds (35% of the total thresholds), the drop in separation reliability was just mild (1.5%, from 0.951 to 0.940), the model fit increased significantly, and multidimensionality was reduced. At the same time, the Spearman's Rho correlation coefficient between the total scores of the original EFA (20 items) and the EFA-R (12 items) was 0.978 (Spearman's rho), which gives an extremely high value of shared variance between the two scales ( $0.978^3=0.956$ ). In synthesis, the loss of information in the revised scale is minimal, precisely as we would expect in the case of item/thresholds multidimensional to the main latent variable and, hence, do not contribute to the validity and the information of the total score of the EFA-R.

We also hypothesized the clinical reasons for these violations of the deleted items. The explanation already given for EFA01 (vegetative stability) can also be applied to EFA02 (Wakefulness/fatigue) and EFA03 (Positioning). The achievement of a stable sleep-wake rhythm, the resistance to strains exceeding 10 minutes, and the possibility of maintaining all positions in the bed can be considered more preconditions for early functional abilities. They may contribute to describing, together with EFA01, a different construct like the “stability of clinical parameters.” EFA08 (Mimic) could represent one of the first signs of wakefulness before the initial communication. EFA09 (Tonus) describes tonus modifications of the extremities, which does not seem to express the recovery of early functional abilities. Another deleted item, EFA11 (Trunk control/sitting), is defined in the scale as the reconditioning to the sitting position, together with the possibility to maintain this position independently with an active and symmetrical trunk erection. Once again, this item probably quantifies different variables like the cardiorespiratory reconditioning to the effort and the trunk control, connected

but separated constructs than functional abilities. Finally, a similar thought can be made for EFA13 (Standing) and EFA15 (Locomotion/mobility in the wheelchair). Indeed, other constructs described by these items are static and dynamic “balance” and “mobility,” which probably constitute higher-level activities and are currently quantified by functional independence scales.

The fact that the described items cannot contribute to measuring the main construct does not imply that they are neither clinically useful nor usable. For instance, in our clinical contexts, we regularly use the EFA item 11 to assess trunk control/sitting. By practice, we have observed that for scores in this item ranging from 1 to 3, the total score of the Trunk Control Test (TCT) is always 0. In contrast, a score of 5 is always associated with TCT scores  $\geq 12$ , while a score of 4 is associated with a TCT score of  $\leq 12$ . Given the overlap between the two scales, we use EFA11 as a guideline on when to start assessing trunk control with the TCT. This highlights that the deleted items may still provide valuable clinical information if used as single-item scales.

In seven of the 13 remaining items, disordered thresholds were resolved following the “classical” approach of collapsing adjacent score categories to obtain an ordered threshold structure for all items. Besides, fit to the model and separation reliability improved further,<sup>45</sup> although the degree of multidimensionality remained unchanged.

We found widespread violations of local independence within the item set that were solved by creating two testlets (super-items) between clusters of locally dependent items. Following this, fit to the Rasch model improved further, and multidimensionality was resolved. These violations frequently occur in health outcome scales.<sup>33</sup> They may be linked to multidimensionality, where some variations among responses are accounted for by a different latent variable.<sup>15, 20, 46</sup> In the “facio-oral functions and ADL” cluster, items related to swallowing, tongue movements, and facio-oral stimulation showed local dependence on the head control item. It is well-known in the literature that decreased postural control exacerbates swallowing and oral movement disorders.<sup>38</sup> In turn, this item demonstrated a high correlation with “excretory functions” and the ability to transfer to the toilet seat, the wheelchair, and the realization of self-care tasks. Indeed, these latter activities require good head and trunk postural control. Simultaneously, “eating” requests the integrity of oral functions, and “transfers to the toilet” becomes significant in case of at least a partial integrity of excretory functions. In the other “sensorimotor and cognitive functions” cluster, items that

quantify voluntary movements and tactile information show local dependence on visual, acoustic functions, and communication. This result is not surprising because, in these patients, the cited functions are related and influence the others. They are also measured by CRS-R items, of which the internal construct validity and reliability have already been demonstrated.<sup>13</sup>

Our solution allowed the calibration of an “essentially unidimensional” reduced subset of 12 items (EFA-R), covering all four original conceptual domains with at least one item. The “essential unidimensionality” was defined by the satisfaction of all four described requirements, which represent the evolution of Stout’s initial definition:<sup>9, 10</sup> local independence between item residuals, a correlation between subtests over 90%, a proportion of explained common variance across subtests over 90%, and a significant difference between estimates of person parameters less than 10%. These findings are different but not in opposition to Poulsen *et al.*’s results, which concluded that EFA subscales measure four latent variables and, thus, could not be summarized into a single EFA Total Score.<sup>7</sup> In contrast, we aimed to select a valid content subset of items from the four EFA dimensions that measured the same underlying latent construct (“early functional abilities”).

Besides, the analysis suggested that the final item hierarchy of the 12 items was consistent with the theoretical and expected hierarchy of functional recovery in patients with DOC. The earlier recovered functioning aspects are processing sensory information (EFA16: tactile information, EFA18: acoustic information) and performing basic motor acts (EFA14: voluntary motricity, EFA10: head control). On the other hand, the latest functioning aspects to be recovered are the more complex motor and cognitive abilities (EFA20: self-care ability, EFA12: transfers) and the control of the excretory functions (EFA04). Table IV also showed that all the four subdomains of the original version of EFA were represented in the revised version: vegetative (1 item), facio-oral (3 items), sensorimotor (3 items), perceptual and cognitive functions (5 items).

Although we found uniform DIF for setting (severe brain injury unit vs. rehabilitation unit vs. nursing home) in the cognitive and sensorimotor testlet, we demonstrated its negligible impact on the person estimates. It is well known that DIF could be “artificial” (*i.e.*, just an artifact of the statistical method employed to detect it).<sup>46</sup> Therefore, this result could also be considered an indirect demonstration of the inter-rater reliability of EFA-R. Furthermore, consistent with our findings, Poulsen *et al.* did not find evidence of DIF for sex and age.<sup>7</sup>

In our sample, we calculated the reliability index in the Rasch model context dependently on the person-item distribution using PSI (0.887), which was adequate for the individual patient measurement. A reliability level >0.850 is crucial for clinicians because it is indispensable to perform precise measurements at the subject level. Indeed, this precision level is mandatory for quantifying the variable, the treatment effectiveness, the patient’s improvement over time, and the comparison between different patient levels of the variable. Similarly, Poulsen *et al.* showed high-reliability values for all EFA subscales, although this was calculated using Hammond & Mesbah’s Monte Carlo estimates.<sup>7</sup>

In Poulsen’s work, three of the four EFA subscales (vegetative functions, facio-oral functions, and sensorimotor functions) were somehow off-target, considering that the sample mean ability was lower than the item set mean difficulty.<sup>7</sup> On the contrary, our analysis showed good targeting, with negligible floor and ceiling effects and a mean-person ability close to the mean item difficulty. These differences could be explained considering that we enrolled a composite sample in terms of ability, selecting one observation randomly for each patient collected at any time during patients’ recovery. In this way, we considered patients with DOC and those who emerged from it and measured precisely the different range of “early functional abilities” of the two states.<sup>35</sup>

Based on our item calibration, it is possible to use the designed table to convert the scale raw scores into interval-level estimates of early functional abilities. As the latter satisfy the requirements for interval-level measurement, clinicians and researchers are encouraged to use these linear estimates, which, unlike the total scores, have absorbed all the local dependence causing multidimensionality. In other words, the EFA-R measures satisfy all the requirements for a scientific measurement like those of the physical sciences (*e.g.*, centimeters, kilograms, etc.).<sup>35</sup> Moreover, they are essential for correctly interpreting change scores<sup>45</sup> and the possibility of using parametric statistics (*e.g.*, ANOVA).<sup>20, 42</sup>

Finally, the pattern of intercorrelations between the EFA-R estimates and the total scores of the CRS-R and mFIM™ may be explained in terms of shared item content between the three scales. In particular, the lower correlation between CRS-R and mFIM™ could be interpreted considering that the two scales do not share specific items. On the other hand, EFA-R and CRS-R contain items assessing oro-facial functions, voluntary movements, and sensory and communication functions. In contrast, there

is a similar overlap in item content between EFA-R and mFIM™ (transfers, excretory functions, and self-care tasks). These overlaps explain not only the “very strong” correlation of the EFA-R with the two other scales but provide a substantive explanation for the potential of the EFA-R to overcome both the ceiling and floor effects of the CRS-R and the mFIM™, as shown in Figure 2.

### Limitations of the study

In this study, the number of available observations was sufficient for a large calibration sample but not for a confirmatory sample analysis, which would have further minimized the risk of capitalizing on chance concerning the fit to the model. As shown in another paper,<sup>13</sup> this is attributable to the relative rarity of UWS, MCS, and emergency from MCS conditions.<sup>47</sup> Given this sample size limitation, it would be necessary to replicate these findings in a larger sample.<sup>10</sup> Further observational studies are also needed to test other external validity types, such as the predictive validity of EFA-R estimates on the CRS-R and mFIM™ total scores and the development of usability criteria (*e.g.*, “transitional” cutoff between CRS-R and EFA-R and between EFA-R and mFIM™).

### Conclusions

The EFA-R is an essentially unidimensional subset of 12 items from the Italian version of the EFA with adequate internal construct validity and sufficient reliability for individual patient measurement under the RMT framework. Furthermore, the possibility of converting EFA-R total scores into linear estimates of early functional abilities, which represent scientific measurements of the underlying construct like the physical sciences, allows clinicians and researchers to interpret change scores correctly and use parametric statistics. Our results further support the hypothesis purported by Poulsen’s<sup>7</sup> and other authors<sup>3-6</sup> that the EFA-R may provide “a measurement bridge” between DOC and functional independence scales for patients with severe acquired brain injury.

### References

1. Alvsåker K, Walther SM, Kleffeldgård I, Mongs M, Drægebø RA, Keller A. Inter-rater reliability of the early functional abilities scale. *J Rehabil Med* 2011;43:892–9.
2. Giacino JT, Kalmar K, Whyte J. The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. *Arch Phys Med Rehabil* 2004;85:2020–9.
3. Heck G, Steiger-Bächler G, Schmidt T. Early Functional Abilities

(EFA) – eine Skala zur Evaluation von Behandlungsverläufen in der neurologischen Frührehabilitation. *Neurol Rehabil* 2000;6:125–33.

4. Pedersen AR, Stubbs PW, Nielsen JF. Reducing redundant testing using the Functional Independence Measure and Early Functional Abilities scale during rehabilitation in patients with brain injury. *Brain Inj* 2018;32:1090–5.
5. Hankemeier A, Rollnik JD. The Early Functional Abilities (EFA) scale to assess neurological and neurosurgical early rehabilitation patients. *BMC Neurol* 2015;15:207.
6. 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;36:2278–81.
7. Poulsen I, Kreiner S, Engberg AW. Validation of the Early Functional Abilities scale: an assessment of four dimensions in early recovery after traumatic brain injury. *J Rehabil Med* 2018;50:165–72.
8. Stout WF. A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika* 1987;52:589–617.
9. Stout WF. A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika* 1990;55:293–325.
10. Mallinson T. Rasch analysis of repeated measures. *Rasch Measure Trans* 2011:1317.
11. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013;310:2191–4.
12. Koller M, Aaronson NK, Blazeby J, Bottomley A, Dewolf L, Fayers P, *et al.*; EORTC Quality of Life Group. Translation procedures for standardised quality of life questionnaires: The European Organisation for Research and Treatment of Cancer (EORTC) approach. *Eur J Cancer* 2007;43:1810–20.
13. La Porta F, Caselli S, Ianes AB, Cameli O, Lino M, Piperno R, *et al.* Can we scientifically and reliably measure the level of consciousness in vegetative and minimally conscious States? Rasch analysis of the coma recovery scale-revised. *Arch Phys Med Rehabil* 2013;94:527–535.e1.
14. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. *Adv Clin Rehabil* 1987;1:6–18.
15. Marais I. Local Dependence. In: Christensen KB, Kreiner S, Mesbah M, editors. *Rasch Models in Health*. London, Hoboken, NJ: ISTE Ltd and John Wiley & Sons, Inc.; 2013.
16. Christensen KB, Makransky G, Horton M. Critical Values for Yen’s Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Appl Psychol Meas* 2017;41:178–94.
17. Andrich D. Components of Variance of Scales With a Bifactor Subscale Structure From Two Calculations of  $\alpha$ . *Educ Meas* 2016;35:25–30.
18. Rodriguez A, Reise SP, Haviland MG. Evaluating bifactor models: calculating and interpreting statistical indices. *Psychol Methods* 2016;21:137–50.
19. Stucky BD, Edelen MO, Vaughan CA, Tucker JS, Butler J. The psychometric development and initial validation of the DCI-A short form for adolescent therapeutic community treatment process. *J Subst Abuse Treat* 2014;46:516–21.
20. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007;57:1358–62.
21. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002;3:205–31.
22. Hardouin JB. MSP: Stata module to perform the Mokken Scale Procedure. *Stata J* 2004;1–21.
23. Sijtsma K, van der Ark LA. A tutorial on how to do a Mokken scale



analysis on your test and questionnaire data. *Br J Math Stat Psychol* 2017;70:137–58.

24. Revicki DA, Chen W, Tucker CA. Developing item banks for patient-reported health outcomes. In: Reise PS, Revicki DA, editors. *Handbook of Item Response Theory Modeling: applications to typical performance assessments*. New York, NY: Routledge; 2014.

25. Lundgren-Nilsson Å, Jonsdottir IH, Ahlborg G Jr, Tennant A. Construct validity of the Psychological General Well Being Index (PGWBI) in a sample of patients undergoing treatment for stress-related exhaustion: a Rasch analysis. *Health Qual Life Outcomes* 2013;11:2.

26. Byrne BM. *Structural equation modeling with AMOS - 2nd edition*. New York, NY: Taylor & Francis; 2010.

27. Muthén LK, Muthén BO. *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén; 2010.

28. Hoyle RH. *Confirmatory factor analysis*. In: Press seSDA, editor. Tinsley H, Brown S, *Handbook of applied multivariate statistics and mathematical modelling*. Cambridge, MA: Academic Press; 2000.

29. La Porta F, Caselli S, Susassi S, Cavallini P, Tennant A, Franceschini M. Is the Berg Balance Scale an internally valid and reliable measure of balance across different etiologies in neurorehabilitation? A revisited Rasch analysis study. *Arch Phys Med Rehabil* 2012;93:1209–16.

30. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6:1–55.

31. Samejima F. *Graded Response Model*. In: van der Linden WJ, Hambleton RK, editors. *Handbook of Modern Item Response Theory*. New York, NY: Springer; 1997.

32. Andrich D. *Rasch models for measurement*. London: Sage Publications; 1988.

33. Pellicciari L, Piscitelli D, Caselli S, *et al*. A Rasch analysis of the Conley Scale in patients admitted to a general hospital. *Disabil Rehabil* 2019;41:2807–16.

34. La Porta F, Giordano A, Caselli S, Foti C, Franchignoni F. Is the Berg Balance Scale an effective tool for the measurement of early postural control impairments in patients with Parkinson's disease? Evidence from Rasch analysis. *Eur J Phys Rehabil Med* 2015;51:705–16.

35. Hobart J, Cano S. Improving the evaluation of therapeutic interven-

tions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009;13:iii, ix–x, 1–177.

36. Kreiner S. The Rasch model for dichotomous items. In: Christensen KB, Kreiner S, Mesbah M, eds. *Rasch Models in Health*. London UK, Hoboken NJ: ISTE Ltd and John Wiley & Sons, Inc.; 2013.

37. Pellicciari L, Piscitelli D, Basagni B, De Tanti A, Algeri L, Caselli S, *et al*. 'Less is more': validation with Rasch analysis of five short-forms for the Brain Injury Rehabilitation Trust Personality Questionnaires (BIRT-PQs). *Brain Inj* 2020;34:1741–55.

38. Redstone F, West JF. The importance of postural control for feeding. *Pediatr Nurs* 2004;30:97–100.

39. Fisher WPJ. Rating Scale Instrument Quality Criteria. *Rasch Measurement Transactions* 2007;21:1095.

40. Kreiner S, Christensen KB. Person parameter estimation and measurement in Rasch Models. In: Christensen KB, Kreiner S, Mesbah M, editors. *Rasch Models in Health*. London UK, Hoboken NJ: ISTE Ltd and John Wiley & Sons, Inc.; 2013.

41. Mueller MK, Item Fit Statistics S. in *Common Software for Rasch Analysis*. Research Report 15-06, Department of Biostatistics, University of Copenhagen; 2015.

42. Maritz R, Tennant A, Fellinghauer C, Stucki G, Proding B. The Functional Independence Measure 18-item version can be reported as a unidimensional interval-scaled metric: internal construct validity revisited. *J Rehabil Med* 2019;51:193–200.

43. Linacre JM, Heinemann AW, Wright BD, Granger CV, Hamilton BB. The structure and stability of the Functional Independence Measure. *Arch Phys Med Rehabil* 1994;75:127–32.

44. Pallant J. *SPSS Survival Manual*. Second Edition. Maidenhead, UK: Open University Press; 2005.

45. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 2007;46:1–18.

46. Andrich D, Sheridan BS, Luo G. RUMM2030: Rasch Unidimensional Models for Measurement. Perth: RUMM Laboratory.

47. Wilson SL, Powell GE, Brock D, Thwaites H. Behavioural differences between patients who emerged from vegetative state and those who did not. *Brain Inj* 1996;10:509–16.

*Conflicts of interest.*—The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

*Authors' contributions.*—All authors read and approved the final version of the manuscript.

*Acknowledgements.*—Participating Investigator: the authors acknowledge all professionals working at the Rehabilitation Unit of the Civil Hospital of Baggi-ovara (Azienda Ospedaliero-Universitaria di Modena), at the Neurorehabilitation Unit "Casa dei Risvegli" (IRCCS Istituto Scienze Neurologiche Bologna), and at Segesta S.p.A. (Korian Group) who collected data.

*History.*—Article first published online: September 28, 2022. - Manuscript accepted: September 26, 2022. - Manuscript revised: September 2, 2022. - Manuscript received: March 16, 2022.

*Supplementary data.*—For supplementary materials, please see the HTML version of this article at [www.minervamedica.it](http://www.minervamedica.it)