'Less is more': validation with Rasch analysis of five short-forms for the Brain Injury Rehabilitation Trust Personality Questionnaires (BIRT-PQs)

Leonardo Pellicciari , Daniele Piscitelli , Benedetta Basagni , Antonio De Tanti , Lorella Algeri , Serena Caselli , Maria Paola Ciurli , Jessica Conforti , Anna Estraneo , Pasquale Moretta , Maria Grazia Gambini , Maria Grazia Inzaghi , Gianfranco Lamberti , Mauro Mancuso , Maria Luisa Rinaldesi , Matteo Sozzi , Laura Abbruzzese , Marina Zettin & Fabio La Porta

View supplementary material

Published online: 12 Nov 2020.

Submit your article to this journal

Article views: 128

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

'Less is more': validation with Rasch analysis of five short-forms for the Brain Injury Rehabilitation Trust Personality Questionnaires (BIRT-PQs)

Leonardo Pellicciari [a], Daniele Piscitelli [b,c,d], Benedetta Basagni [e], Antonio De Tanti [e], Lorella Algeri[f], Serena Caselli [g,h], Maria Paola Ciurli [i], Jessica Conforti[e], Anna Estraneo [j,k], Pasquale Moretta[l], Maria Grazia Gambini [m], Maria Grazia Inzaghi[n], Gianfranco Lamberti [o], Mauro Mancuso [p,q], Maria Luisa Rinaldesi[r], Matteo Sozzi [s], Laura Abbruzzese [q], Marina Zettin[t,u], and Fabio La Porta [v]

aNeurorehabilitation Research Laboratory, IRCCS San Raffaele Pisana, Rome, Italy; bSchool of Physical and Occupational Therapy, McGill University, Montreal, Canada; cSchool of Medicine and Surgery, University of Milano Bicocca, Milan, Italy; dSchool of Physical Therapy and Athletic Training, Pacific University, Hillsboro, OR, USA; eCardinal Ferrari Centre, S. Stefano Rehabilitation, Fontanellato (PR), Italy; fUOC Psicologia, ASST Papa Giovanni XXIII, Bergamo, Italy; gUnità Operativa di Medicina Riabilitativa, Dipartimento di Neuroscienze, Azienda Ospedaliero-Universitaria di Modena, Modena, Italy; hScuola di Dottorato in Sanità Pubblica, Università degli Studi di Milano-Bicocca, Milano, Italy; iUnità Post-Coma, IRCCS Fondazione Santa Lucia, Rome, Italy; jSevere Brain Injury Department, IRCCS Don Gnocchi Institute, Florence, Italy; kNeurology Unit, Santa Maria della Pietà General Hospital, Nola, Italy; lMaugeri Scientific and Clinical Institutes, IRCCS, Department of Neurorehabilitation, Institute of Telese Terme, Telese Terme, Italy; mDipartimento di Riabilitazione, Ospedale Sacro Cuore- Don Calabria, Verona, Italy; nLaboratorio di Neuropsicologia, Istituto Clinico Quarenghi, Bergamo, Italy; oDepartment of Rehabilitative Medicine, AUSL Piacenza, Piacenza, Italy; pDepartment of Rehabilitation, National Health Service South-Est Tuscany, Grosseto, Italy; qResearch Centre, Tuscany Rehabilitation Clinic, Montevarchi, Italy; rSanto Stefano Riabilitazione, Istituto di Riabilitazione Santo Stefano, Porto Potenza Picena, Italy; sNeurology Unit, "A. Manzoni" Hospital, Lecco, Italy; tCentro Puzzle, Turin, Italy; uDipartimento di Psicologia, Università di Torino, Turin, Italy; vIRCCS Istituto delle Scienze Neurologiche di Bologna, Neurorehabilitation Unit, Bologna, Italy

ABSTRACT

Background: Previous analyses demonstrated a lack of unidimensionality, item redundancy, and substantial administrative burden for the Brain Injury Rehabilitation Trust Personality Questionnaires (BIRT-PQs).

Objective: To use Rasch Analysis to calibrate five short-forms of the BIRT-PQs, satisfying the Rasch model requirements.

Methods: BIRT-PQs data from 154 patients with severe Acquired Brain Injury (s-ABI) and their caregivers (total sample = 308) underwent Rasch analysis to examine their internal construct validity and reliability according to the Rasch model.

Results: The base Rasch analyses did not show sufficient internal construct validity according to the Rasch model for all five BIRT-PQs. After rescoring 18 items, and deleting 75 of 150 items, adequate internal construct validity was achieved for all five BIRT-PQs short forms (model chi-square p-values ranging from 0.0053 to 0.6675), with reliability values compatible with individual measurements.

Conclusions: After extensive modifications, including a 48% reduction of the item load, we obtained five short forms of the BIRT-PQs satisfying the strict measurement requirements of the Rasch model. The ordinal-to-interval measurement conversion tables allow measuring on the same metric the perception of the neurobehavioral disability for both patients with s-ABI and their caregivers.

Introduction

Behavioral disorders are frequent in acquired brain injuries (ABI), both of traumatic and non-traumatic origin (TBI and non-TBI, respectively) (1,2). The prevalence of such disorders ranges between 30% and 60% of individuals with TBI and non-TBI[, respectively. Several authors (1,2) referred to these behavioral disorders as neurobehavioral or personality changes (or neurobehavioral disability) affecting the activity and social participation of individuals post-ABI. Manifestations of these changes might range from apathy, impulsivity, and extreme sensitivity to criticism, to aggression and difficulties building and/or maintaining a reciprocal relationship. Furthermore, neurobehavioral disability has been strongly associated with poor health outcomes on the quality of life of subjects post-ABI and

their caregivers (3,4,5). Interestingly, neurobehavioral changes due to ABI were better predictors of subjective family distress than disease severity or cognitive impairments (6). As a result, the ongoing social and family burden of care remains high.

Despite their clinical relevance for the health status, only a few clinical measures have been developed to assess neurobehavioral changes after ABI (7,8). Oddy et al (9) proposed the Brain Injury Rehabilitation Trust Personality Questionnaires (BIRT-PQs), a set of novel patient-reported outcome measures aimed at investigating neurobehavioral disability occurring in individuals with ABI. The BIRT-PQs include five separate questionnaires (each available in parallel forms for patient and caregiver), for a total of 150 items, which assess various areas of neurobehavioral changes following severe ABI (s-ABI).

Although preliminary analyses supported the internal consistency, test-retest reliability, and the external construct validity of the five BIRT-PQs (2,9,10), we recently conducted a multicenter psychometric study on the Italian version of the BIRT-PQs (11), which somehow challenged this initial evidence. In particular, our study focused on the internal construct validity (ICV) of the BIRT-PQs on a pooled sample of data (N = 308) from the patients and their respective caregivers within the Classical Test Theory (CTT) psychometric framework. ICV aims to establish the validity of the total score of a scale by comparing the psychometric performance of its items to that predicted by a measurement model (12,13). Notably, an internal consistency analysis revealed that, despite good overall internal consistency values for each scale, several items within each questionnaire contributed much less than expected to the total score.

A Confirmatory Factor Analysis (CFA) confirmed this finding showing the misfit of the data of each scale to a one-factor model, which suggests a lack of unidimensionality for the total scores. Furthermore, the baseline CFAs showed another severe violation of ICV, represented by the presence of local dependency (LD) between 110 pairs of items, which indicated item redundancy. Indeed, after accounting for this LD, the fit to a one-factor CFA model improved significantly, thus suggesting that item redundancy could be an essential source of the misfit. Beyond multidimensionality and LD, another critical source of misfit for ICV is the violation of multi-group invariance. It occurs when one or more items give different success rates for two or more groups at the same ability level (14) Unfortunately, we were unable to test for possible violations of multi-group invariance by etiology and respondent, given the larger sample size needed for this kind of analysis within the CTT framework (8).

Overall, the above results paved the way for a more in-depth assessment of the ICV of the BIRT-PQs within the Rasch Measurement Theory (RMT) framework. Rasch analysis (15,16,17) is the statistical process of testing whether the responses to the items of a scale or questionnaire fit the measurement and psychometric requirements of a family of mathematical models named the 'Rasch models' (18) after the Danish mathematician Georg Rasch (1901–1980). In fact, the RMT allows detailed testing and adjusting for all violations of ICV requirements (12) including violations of unidimensionality (15,19,20), local independence (19,21), monotonicity (15,19), and multi-group invariance (also known as Differential Item Functioning [DIF]) for different factors, such as gender, age, etiology, and respondent (15,16,19,22). Within the RMT, the analysis of DIF is performed, item by item, on the whole dataset, thus allowing to test for DIF by respondents (patient vs. caregiver) on the pooled data and circumventing the sample size limitations imposed by the CTT methods.

Furthermore, the RMT offers a variety of techniques to adjust for the ICV requirements violations, including the possibility to delete misfitting or redundant items. This is desirable for the BIRT-PQs because we showed not only item redundancy within the CTT analyses (11), but we also demonstrated that the BIRT-PQs were demanding instruments in terms of administration time, both for patients and caregivers. Beyond

ICV, it is possible to assess also the separation reliability and the targeting of the scale to the sample (15,23,24). Finally, the RMT allows testing for item invariance (homogeneity), which implies that the difficulty order of the items is independent of the ability of the persons affirming them (15,17,24). Invariance constitutes a salient feature of all measurement processes within the physical sciences. Indeed, the Rasch models are considered a stochastic operationalization of the formal axioms of Additive Conjoint Measurement, a general measurement theory (25,26). Therefore, within the RMT, when data fit the model's requirements, item difficulties and person abilities will be estimated along the same measurement continuum. Thus, given some specific statistical properties acquired by the total score (i.e., specific objectivity and sufficiency), it will be possible to transform it from an ordinal to a linear interval-level scale of neurobehavioral change, whose unit of measurement is the logit (15,17,27,28).

Therefore, this study aimed: 1) to assess the ICV of the five BIRT-PQs within the RMT framework; 2) to address any threats to the ICV of the BIRT-PQs, using primarily item reduction techniques, to develop shorter forms of the BIRT-PQ; 3) to ensure that these shorter forms have adequate ICV and sufficient precision for individual person measurement.

Methods

Subjects and setting

Full details on the study methodology, including the enrollment procedures and the setting, were provided elsewhere (11). Briefly, eleven Italian neurorehabilitation centers for patients with brain injury enrolled consecutively the participants and their caregivers from April 2016 to December 2017. Individuals with brain injury met the inclusion criteria if aged between 18–70 years, suffered from a severe ABI characterized at the onset by lack of consciousness (Glasgow Coma Scale ≤8) lasting more than 24 hours, had a Level of Cognitive Functioning score ≥7 at the time of enrollment, and were independent before the s-ABI (i.e., modified Barthel Index = 100) (11). Exclusion criteria were the presence of severe aphasia and previous history of neurological and psychiatric disorders. The local Ethical Committees of the participating centers approved the study, which was carried out following the principles outlined in the Helsinki declaration (29). Participants and their respective caregivers gave their written informed consent to take part in the study.

Outcome measures

The BIRT-PQs (9,10,30) are a set of patient-reported outcome measures, making up five separate questionnaires: motivation (BMQ, 34 items), regulation of emotions (BREQ, 32 items), social cognition (BSCQ, 28 items), disinhibition (BDQ, 24 items), and impulsivity (BIQ, 32 items). The BIRT-PQs are available in two versions: a self-rated patient version and a caregiver-rated version. The former casts light on the patient's perception of his/her own aspects of neurobehavioral changes, whereas the latter considers the caregiver's perspective on the same elements of their relative. Each item is rated

on a 4-level Likert scale, ranging from 1 (always) to 4 (never). The total score of each questionnaire is computed by summing the scores of each item: BMQ: ranged from 34 to 136 points; BREQ from 32 to 128; BSCQ from 28 to 112; BDQ from 24 to 96; and BIQ from 32 to 128. Higher total scores indicate higher degrees of neurobehavioral disability. The Italian versions of the BIRT-PQs (31)were administered to all patients and their caregivers before starting their physical therapy program.

Assessment of the conceptual content of the BIRT-PQ

This procedure was performed to facilitate the interpretation of the findings of the Rasch analyses and to guide the scale modification procedures. Two authors (FLP and BB) conducted an appraisal of the conceptual content of each BIRT-PQ separately by linking the items of each questionnaire to the conceptual categories suggested by Hyde (30). Linking inconsistencies and difficulties were subsequently resolved in a joint session, also involving three of the other authors (ADT, SC, LP), where conceptual categories deemed to be too similar were collapsed to find the minimum set of theoretical facets represented by each questionnaire.

Rasch analysis

Rasch analysis has been described in detail elsewhere (13,32,33,34,35). A full description of the methods used to assess the measurement quality of a scale within the current Rasch analysis is available in Appendix 1 (supplemental digital content). The choice of the appropriate polytomous Rasch model (rating scale vs. partial credit) was made based on the results of a Fisher's likelihood ratio test run for each BIRT-PQs (36). To interpret the Rasch analysis outputs, we reported the following summary statistics:

- **Fitness to the Rasch model** relates to the stochastically invariant ordering of the items. It was summarized using the mean and the standard deviation (SD) of the item and person fit residuals, as well as a summary chi-square interaction statistic. We considered an adequate fit to the model achieved when the SD of the item and the person fit residuals was ≤1.4 (37), and the summary chi-square was not significant (i.e., values were above the Bonferroni correction), thus indicating no deviation from the model's expectations (33,38).
- **ICV requirements**. We summarized the findings using the following summary statistics:
  a. Unidimensionality, which prescribes that all items in a scale must measure a single underlying construct (34,39). We tested this critical ICV requirement employing a paired t-test conducted on separate estimates for each subject (derived from subsets of items identified by principal component analysis of the residuals) (20). We considered strict unidimensionality achieved when both the proportion of significant tests (PST) and the lower bound of the binomial confidence interval for proportions (BCI) were below 5%. In contrast, unidimensionality was considered acceptable when only the BCI was <5%.

  b. Monotonicity, which prescribes that the probability of endorsing an item or response option indicative of higher neurobehavioral disability should increase with the increase of the underlying latent trait. This requirement was summarized as the percentage of items with disordered thresholds (T-DT), expecting a value of 0% for adequate monotonicity.
  c. Local independence prescribes that all the variation among responses to an item is accounted for by the person's ability only and, therefore, for the same value of ability, there is no further systematic relationship among responses. We considered items to be locally independent if their residual correlation was above a Local Dependency Relative Cutoff (LDRC), calculated by adding 0.2 to the average of residual correlations, after having removed the association of each item to itself, equal to 1 (34,40). We summed all the correlation coefficients of the residuals above the LDRC to obtain a total value of LD (T-DL), where 0 indicates the complete absence of LD.
  d. Absence of DIF prescribes that an item must be invariant also across relevant subgroups (or person factors), such as gender or age. In this case, different groups of persons, with equal levels of the underlying characteristics within a person factor, respond in the same manner regardless of their group membership. We tested the presence of DIF with a two-way ANOVA for each item, where scores are compared across each level of the person factor and different ability levels, as summarized by the class intervals. DIF is present when the p-values of ANOVA are significant below the Bonferroni-correction (41). We summarized the amount of DIF (T-DIF) by obtaining the absolute value of the base-ten logarithm of the sum of all significant p-values across all items and all person factors. T-DIF values range from zero to infinite, where zero indicated no DIF. We tested the following person factors within the DIF analysis: age (under 48 years vs. over 48 years, based on the median age calculated on the total sample); gender (male vs. female); educational level (under 8 years vs. over 8 years); etiology (TBI vs. no TBI); time since lesion (above 18 months vs. below 18 months, based on the median time since lesion computed on all patients); and respondent (patients vs. caregivers).

- **Reliability and targeting** were summarized as follows:
  a. Targeting, which indicates how well the measurement range of the scale matches the distribution of the calibrating sample (23), was assessed in terms of floor and ceiling effects and Targeting Index (TI). We established that targeting was good and fair for ranges of TI [−1, +1] and [−2, +2], respectively.
  b. Separation reliability, which is the ability of the scale to separate persons effectively based on their level of ability, was expressed as Person Separation Index (PSI), Cronbach's alpha (α), number of statistically Distinct Levels of Performance Ability (DLPA) (35), and Distribution-Independent Person Separation

Index (DI-PSI) (42). We established that PSI or DI-PSI values ≥0.85 and ≥0.70<0.84 were sufficient for individual person and group-level measurements, respectively(43–45).

When the data did not fit the Rasch model (as it is often the case), the questionnaires were progressively modified to adjust for the violations of the ICV requirements. This process, which we undertook iteratively, was based on post-hoc item modifications that were:

(a) Structural, where the structure of the scale was actively modified, either because of rescoring or deleting an item. These modifications affected the total score range.
(b) Statistical, where the structure of the scale was unmodified (the total score range remained unchanged), but these adjustments (testlets creation and item splitting) affected mainly the conversion of the total score into interval level estimates of ability.

Considering the second aim of the present study (i.e., to ease questionnaires administration), we gave priority to structural modifications in case of lack of fitness to the model, by performing:

- Item rescoring, which is the collapsing of adjacent response categories of the same item to resolve the violations of monotonicity. Published guidelines were followed (46), and the rescoring pattern was carried out to maximize statistical indexes and clinical meaning (13,47);
- Item deletion, where one item at a time was deleted according to the following criteria: the presence of LD with one or more items, misfit to the model requirements, presence of DIF, lower number of response categories, and clinical meaning (13,48);

Should the above approaches failed, item grouping for creating 'testlets' (49) and item splitting (22) would be performed to account for any remaining violation of the local independence and absence of uniform-DIF requirements, respectively. Fitness to the Rasch model, ICV requirements, reliability, and targeting were all assessed for the original scale (base analysis) and, then, after each scale modification, to ascertain whether adequate model fit was achieved. This process was repeated cyclically until no further changes were needed and/or possible.

Statistical notes, software and sample size issues
Rasch analysis was run with RUMM2030 software (50). It was estimated that a sample size of about 300 subjects would be sufficient to estimate item difficulty with α of 0.01 to <±0.5 logits, irrespective of the targeting of persons to items (51). A significance value of 0.05 was used throughout and corrected for the number of tests by Bonferroni correction (52). We used the RUMM Logbook™ (53), an *ad hoc* Excel 2007™ application developed using Microsoft Visual Basic™ macros to facilitate the interpretation of the results of each Rasch analysis. A free copy of this application is available from the corresponding author upon request.

Results

Sample characteristics
One-hundred and fifty-four subjects with s-ABI (mean age: 41.9 years; SD: 14.4; 68.8% males) and their 154 respective caregivers (mean age: 52.1 years; SD: 12.3; 68.8% females) were enrolled in this study. The main demographic and clinical characteristics of participants are reported in Table 1. Further detailed characteristics are available elsewhere (11).

Rasch analysis summary
Within the present study, the Rasch analyses were based on the partial credit parameterization of the Rasch model, as Fisher's likelihood ratio tests performed on each scale were all significant, thus rejecting the simpler rating scale model. Briefly, the base Rasch analyses failed to demonstrate adequate fit to the Rasch model and satisfaction of all the ICV requirements for all five BIRT-PQs, including LD, item misfit, disordered thresholds (DT), and presence of DIF.

However, after extensive structural modifications, which reduced the total item load of the five BIRT-PQs from 150 to 78 items, we were able to achieve adequate fit to the Rasch model and the satisfaction of the ICV requirements with reliability levels compatible with person measurement for all five BIRT-PQ. Table 2 displays a summary of the Rasch analysis of each BIRT-PQ.

The specific tables detailing the item fit statistics and the scoring model for the final solution of each BIRT-PQs are presented in Appendix 2 (supplemental digital content). Besides, the targeting graphs of each final solution are

Table 1. Main demographic and clinical characteristics of the sample (N = 308).

Variable	Patients (N = 154)	Caregivers (N = 154)
Age (*years*)	41.9 ± 14.4*	52.1 ± 12.3*
Gender		
Males	68.8%	31.2%
Females	31.2%	57.8%
Missing	0.0%	11.0%
Education		
Primary	2.6%	5.8%
Secondary	31.2%	30.5%
High	48.7%	37.7%
Degree	16.9%	13.6%
Post-degree	0.0%	1.3%
Missing	0.6%	11.0%
Etiology		
Traumatic brain injury	61.0%	-
Intracerebral hemorrhage	13.0%	-
Subarachnoid hemorrhage	9.7%	-
Ischemic stroke	6.5%	-
Anoxia	4.5%	-
Other	5.2%	-
Time since lesion (*months*)	26.3 ± 19.6*	-
BIRT-PQs[a]		
Motivation (range 34–136)	63.0 [49.0, 84.0]†	68.0 [48.0, 98.1]†
Emotional regulation (range 32–138)	50.0 [38.0, 72.8]†	51.0 [40.0, 78.0]†
Social cognition (range 28–112)	46.0 [38.0, 62.8]†	45.0 [34.0, 69.0]†
Disinhibition (range 24–96)	42.0 [34.0, 52.7]†	43.0 [34.0, 57.0]†
Impulsivity (range 32–128)	56.0 [45.0, 73.8]†	59.0 [44.0, 78.2]†

Abbreviations: BIRT-PQ, Brain Injury Rehabilitation Trust; N, number of cases.
Notes:
All data are reported as a percentage, except * which indicates mean ± SD, and †, which indicates median and the interpercentile range [10P, 90P].
[a]higher values indicate a higher level of behavioral disturbance.

displayed in Figure 1. Also, we reported in Table 3 the raw score to interval-level measure conversion table for the final analyses of each BIRT-PQ. The patient and relative versions of BIRT-PQs short-forms are presented in Appendix 3 and 4, respectively. Finally, the Italian versions (patient and caregiver) of the BIRT-PQs short-forms are available at https://sstefano.it/birt-short-version.

Figure 2 provides the BIRT-PQ patient-caregiver profiles for four different cases in the sample. In each chart, after normalizing the measurement range of all questionnaires into a 0–100 rescaled logit range, the patient's rating was plotted against the caregiver's rating. In this way, it is possible to appraise both substantial disagreements between the two measures and the hierarchy of the severity of the neurobehavioural disability captured by the five BIRT-PQ.

BIRT motivation questionnaire (BMQ)

We identified five main conceptual facets of the BMQ: affective-emotional (including anhedonia, hopelessness, indifference, and lethargy), indecision/lack of ideas, difficulties to initiate a task, lack of organization, and distractability/perseverance.

As shown in Table 2, the base analysis showed that the original 34-item BMQ failed to satisfy the requirements of stochastic invariance ($\chi^2_{136} = 310.0$; p ≤ 0.001) and monotonicity as two items showed DT, local independence (26 items had their residual correlation above the LDRC, here set at 0.171), unidimensionality (lower BCI = 11.0%), and absence of DIF (BMQ29, BMQ21, BMQ33 displayed uniform DIF). Furthermore, five items (BMQ01, BMQ06, BMQ13, BMQ18, BMQ32) were underfitting the model (item fit residuals ranging from 2.909 to 4.186).

After rescoring two items with DT, we dealt with clusters of locally dependent items. As expected, we found the higher residual correlations, indicative of LD, within the conceptual facets identified by the content analysis. For instance, the item 'I feel energetic' (BMQ30) was locally dependent with 'I achieve my goals' (BMQ21), 'I am an enthusiastic person' (BMQ24), 'I enjoy life' (BMQ29), 'I am good at making new friends' (BMQ32), and 'I have a lot of gets up and go' (BMQ11), which were all linked to the 'affective-emotional' facet. Some items were subsequently deleted because misfitting, as 'I am good at making new friends' (BMQ32) and 'I plan my week and make arrangements for things to do' (BMQ06), possibly because these items may be influenced by other constructs external to motivation, such as social abilities and executive functions, respectively. Another item, 'I enjoy life,' was deleted because displaying uniform-DIF by the respondent, as patients were more likely to affirm higher scores (indicative of higher levels of perceived motivational impairment) than caregivers given the same level of the construct.

After deleting 15 items in total (80% for LD), the final 19-item set (BMQ-SF19) showed adequate fit to the Rasch model ($\chi^2_{76} = 11.2$; $p = .005$; Bonferroni-adjusted p-value = 0.003) and satisfied the requirements of monotonicity, acceptable unidimensionality (PST = 6.2%; lower BCI = 4.6%), and absence of DIF. There was some left-over LD between a pair of items (BMQ22 and BMQ27), which had a residual correlation of 0.175, which was a value slightly above the LDRC, here set at 0.146). Furthermore, BMQ03 showed marginal under-fit to the model (fit residual = 2.833), indicating that the responses to this item were too unpredictable compared to the model's expectations. At the content level, all five facets were equally represented in the final scale. The item hierarchy showed that 'distractability/perseverance' (average item difficulty in logits: −0.438) was the facet associated with the lower levels of motivational impairments. 'Affective/emotional' (−0.044 logits), 'indecision/lack of ideas' (+0.021 logits), and 'difficulty to initiate' (+0.035 logits), were, on average, associated to increasing levels of motivational impairment, whereas 'lack of organization' (average item difficulty in logits: +0.194) was the facet associated to the higher degree of motivational impairment.

Overall, the scale appeared to be off-target (Figure 1a), as the TI was −3.129, indicating that the mean item difficulty was higher than the mean ability of the sample. Both the PSI and the DI-PSI (0.881 and 0.973, respectively) indicated that the scale was adequate for individual person measurement (Table 2). The deleted items, the item fit statistics, and the scoring model for the BMQ-SF19 are reported in Appendix 2a. The total score of the BMQ-SF19 ranged from 0 to 57 points (Table 3).

BIRT emotional regulation questionnaire (BREQ)

Within the BREQ, we identified four main conceptual facets: 'emotional lability/mood swings', 'irritability/lack of emotional control', 'no reasons/cause for the behavior', and 'outburst consequences'.

As reported in Table 2, the initial Rasch analysis showed that the original 32-item set of the BREQ misfitted the model, failing the requirements of the stochastic ordering of the items ($\chi^2_{128} = 564.6$; p ≤ 0.001), monotonicity (eight items had DT), local independence (25 items had residual correlation values higher than the LDRC, here set at 0.174), and presence of DIF by the respondent for one item. However, its unidimensionality was acceptable (lower BCI = 4.6%). Five items (BREQ01, BREQ02, BREQ05, BREQ16, BREQ26) were under-fitting the model (i.e., their response pattern was too unpredictable), whereas six items (BREQ06, BREQ08, BREQ24, BREQ25, BREQ28, BREQ29) over-fitted the model (i.e., their response pattern was too predictable).

After rescoring all items with DT, locally dependent items were dealt with. For instance, 'my mood can change quickly for no reason' (BREQ05) was found to be locally dependent with 'I suddenly feel angry and do not know why' (BREQ26), 'I lose my temper very suddenly without knowing why' (BREQ02), and 'I have sudden mood swings' (BREQ1), where the first three items belonged to the 'no reason/cause for the behavior' facet. As BREQ05 was also overfitting the model (i.e., the responses to it were too predictable), it was deleted to 'free-up' the other locally dependent items. Instead, in another cluster within the 'irritability/lack of emotional control' facet, even though 'I am calm' (BREQ20) was found to be strongly locally dependent with two items ('I'm in control,' BREQ12 and 'I'm relaxed,' BREQ24), it was the one retained. The latter two were deleted because they were also severely misfitting, probably because exploring emotional states that may be less easy to identify

Table 2. Summary of the Rasch analyses for each of the brain injury rehabilitation trust personality questionnaires.

	Analysis description			Item fit residual		Item fit residual		Item-trait interaction			Unidimensionality		Other ICV requirements			Separation reliability					
BIRT-PQs	Analysis name	N/CI	K	Mean	SD	Mean	SD	χ^2_{df}	p	Cutoff[a]	PST (%)[f]	Lower BCI (%)[f]	T-DT[d]	T-LD[e]	T-DIF[g]	PSI	α	DLPA	DI-PSI	SEM[b]	Targeting index[c]
Motivation (BMQ) Base	Base	308/5	34	0.375	**1.619**	−0.197	**1.709**	310.0_{136}	**0.0000**	≥0.0015	**12.7%**	**11.0%**	**5.9%**	**9.718**	**34.3**	0.921	0.930	10	0.990	0.250	**−3.129**
Final	Final	308/5	19	0.302	1.241	−0.266	**1.456**	111.2_{76}	0.0053	≥0.0026	**6.2%**	4.6%	0.0%	**0.350**	0.0	0.881	0.900	6	0.973	0.362	**−2.573**
Emotional Regulation (BREQ) Base	Base	308/5	32	0.024	**2.672**	−0.154	1.394	564.6_{128}	**0.0000**	≥0.0016	**6.2%**	4.6%	**25.0%**	**12.568**	**7.8**	0.908	0.937	9	0.988	0.342	**−4.138**
Final	Final	308/5	17	−0.175	1.300	−0.270	1.043	89.5_{68}	0.0414	≥0.0029	5.4%	3.7%	0.0%	0.0	0.0	0.850	0.908	6	0.973	0.512	**−3.405**
Social Cognition (BSCQ) Base	Base	308/5	28	0.001	**1.937**	−0.270	**1.471**	264.6_{112}	**0.0000**	≥0.0018	**11.4%**	**9.7%**	**35.7%**	**9.410**	**12.2**	0.887	0.887	8	0.985	0.317	**−3.918**
Final	Final	308/5	13	−0.018	0.864	−0.252	1.046	47.1_{52}	0.6675	≥0.0038	5.0%	3.3%	0.0%	0.0	0.0	**0.783**	**0.817**	5	0.962	0.516	**−2.682**
Disinhibition (BDQ) Base	Base	308/5	24	0.316	**3.133**	−0.120	1.269	563.3_{96}	**0.0000**	≥0.0020	**6.8%**	**5.2%**	**41.7%**	**6.182**	**78.7**	**0.789**	**0.811**	8	0.985	0.277	**−3.473**
Final	Final	308/5	13	−0.049	1.119	−0.257	1.039	64.5_{52}	0.1136	≥0.0038	4.4%	2.7%	0.0%	0.0	0.0	**0.747**	**0.800**	5	0.962	0.529	**−2.938**
Impulsivity (BIQ) Base	Base	308/5	32	0.138	**2.946**	−0.225	**1.657**	609.1_{128}	**0.0000**	≥0.0016	**11.0%**	**9.4%**	**12.5%**	**14.916**	0.0	0.891	0.904	9	0.988	0.259	**−3.881**
Final	Final	308/5	16	−0.018	1.172	−0.238	1.176	85.8_{64}	0.0357	≥0.0031	5.3%	3.6%	0.0%	0.0	0.0	**0.824**	0.857	6	0.973	0.472	**−3.315**
Recommended value →	-	-	-	-	≤1.4	-	≤1.4	-	-	-	<5.0[g]	Lower BCI<5.0	0%	0	0	≥0.85[h]	≥0.85[h]	-	≥0.85[h]	-	[−2, 2]

Abbreviations: ICV, internal construct validity; BIRT-PQs, Brain Injury Rehabilitation Trust Personality Questionnaires; N/CI, the ratio between sample size and class intervals; K: number of items; SD, standard deviation; χ^2, unconditional chi-square for model fit; df, degrees of freedom; p, Bonferroni-corrected probability value for chi-square; PSI, Person Separation Index; α, Cronbach's alpha; DLPA, distinct levels of performance ability; DI-PSI: distribution-independent person separation index (based on DLPA); SEM, standard error of measurement; T-DT, percentage of items with disordered thresholds; T-LD: total local dependency load; PST, the proportion of significant t-test carried out on the estimates that, within a principal component analysis of residuals, loaded positively and negatively (factor loading >0.30) on the first component; BCI, binomial (95%) confidence interval for proportions of significant t-test; T-DIF, total DIF load.

Notes:

a Bonferroni-corrected p-value, which varies by analysis, and that is used to interpret the corresponding chi-square p-value.

b SEM is calculated with the formula: $SD \times \sqrt{1 - reliability}$, where SD is the person's location standard deviation, and reliability is the PSI with extremes.

c 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, respectively, [−2, +2] SEM the average item measure (set by default at 0 logits).

d 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.

e The T-LD summary statistics is calculated by summing together all the residual correlations values above the local dependency relative cutoff, which is calculated as the mean of all the residual correlations (excluding the correlation of items with themselves) minus 0.2. The values range from zero to infinite, where zero indicates the absence of local dependency.

f Unidimensionality is considered achieved either when PST is <5% or when the lower bound of its BCI is <5%.

g 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.

h A value of ≥.850 suggests a precision of measurement also at the individual level, whereas a value between 0.700 and 0.849 indicates precision only at the group level. All values outside the recommended range are in bold.
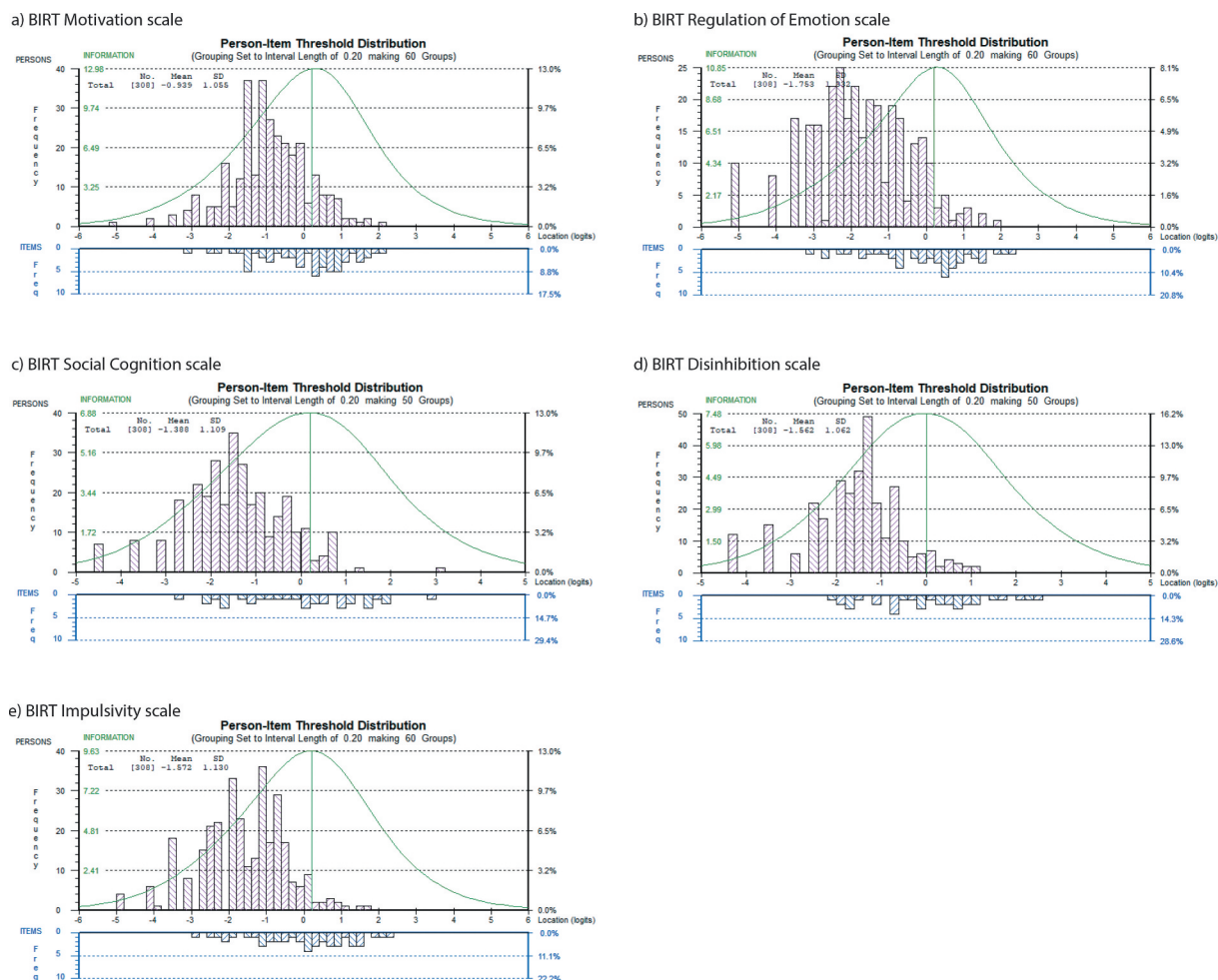
Figure 1. Targeting (person-thresholds distribution) graphs for each Brain Injury Rehabilitation Trust Personality Questionnaires. For each graph, persons (N = 308) and item thresholds are displayed, respectively, in the upper and the lower part of the chart, separated by the logit scale. Grouping set to interval length of 0.20, making 60 groups for Motivation, Emotional Regulation, and Impulsivity questionnaires, and 50 groups for Social Cognition and Disinhibition questionnaires. Abbreviations: BIRT, Brain Injury Rehabilitation Trust; Freq, frequency; No, number; SD, standard deviation.

precisely. On the other hand, 'calmness' (BREQ20) is easy to identify as the opposite of 'rage,' a frequently experienced and problematic emotional state in this population.

After deleting 15 items (14 of which because LD), the final 17-item scale (BREQ-SF17) satisfied all model's requirements, in terms of invariance (χ^2_{68} = 89.5; p = .041), monotonicity, LD, acceptable unidimensionality (PST = 5.4%; lower BCI = 3.7%), and absence of DIF. There were no misfitting items. At the content level, all five conceptual facets were adequately represented. The item hierarchy showed that the facet associated with the lower levels of emotional dysregulation was 'lability/mood swings' (average item difficulty: −0.838 logits). 'Irritability/lack of emotional control' (−0.118 logits) and 'consequences of an outburst' (+0.103 logits) were, on average, associated to increasing levels of emotional regulation impairment, whereas 'no reason/cause for the behavior' (average item difficulty in logits: +0.439) was the facet associated to the higher degree of emotional dysregulation.

A TI of − 3.405 showed that the mean ability of the sample was lower than the mean item difficulty, as also demonstrated by the visual inspection of the targeting graph (Figure 1b). Both the PSI and the DI-PSI (0.850 and 0.973, respectively), suggested precision of measurement at the individual level (Table

3). The item fit statistics and the scoring model for the BREQ-SF17 are reported in Appendix 2b. The total score of the BREQ-SF17 ranged from 0 to 48 points (Table 3).

BIRT social cognition questionnaire (BSCQ)

We identified four main conceptual facets within the BSCQ: 'inability to interpret external cues,' 'theory of mind/lack of empathy,' 'social anxiety,' and 'difficulties in social interaction.'

The base analysis of the original 28-item BSCQ showed misfit to the model expectations in terms of violations of the requirements of stochastic invariance (χ^2_{112} = 264.6; p ≤ 0.001), monotonicity (ten items had DT), local independence (20 items had their residual correlations higher than the LDRC, here set at 0.167), unidimensionality (lower BCI = 9.7%), and absence of DIF (BSCQ08 with a uniform-DIF for etiology). Furthermore, at the item level, BSCQ19 severely under-fitted the model (item fit residual = 8.113) (Table 2).

After rescoring all items with DT, we dealt with locally dependent items. For instance, '*I find it hard to understand what people mean*' (BSCQ02) was found to be locally dependent with '*I find it hard to understand people on the*

Table 3. Raw score to measure estimates conversion table for each brain injury rehabilitation trust personality questionnaires based on the original sample calibrations.

Raw score	Motivation (BMQ-SF19)			Emotional Regulation (BREQ-SF17)			Social Cognition (BSCQ-SF13)			Disinhibition (BDQ-SF13)			Impulsivity (BIQ-SF16)		
	Logit scale	0–100 m-L scale	±95% CI	Logit scale	0–100 m-L scale	±95% CI	Logit scale	0–100 m-L scale	±95% CI	Logit scale	0–100 m-L scale	±95% CI	Logit scale	0–100 m-L scale	±95% CI
0	−5.013	0.0	2.40	−5.030	0.0	2.44	−4.531	0.0	2.40	−4.321	0.0	2.38	−4.882	0.0	2.42
1	−4.165	8.8	1.72	−4.177	8.9	1.76	−3.696	9.2	1.74	−3.503	9.4	1.70	−4.043	8.9	1.74
2	−3.579	15.0	1.37	−3.575	15.2	1.40	−3.106	15.7	1.39	−2.938	15.9	1.34	−3.454	15.1	1.38
3	−3.174	19.2	1.18	−3.149	19.7	1.22	−2.690	20.3	1.21	−2.547	20.4	1.16	−3.042	19.5	1.20
4	−2.859	22.5	1.05	−2.812	23.2	1.10	−2.358	24.0	1.10	−2.241	23.9	1.04	−2.717	23.0	1.08
5	−2.599	25.2	0.97	−2.529	26.2	1.01	−2.077	27.1	1.02	−1.986	26.8	0.97	−2.444	25.9	1.00
6	−2.375	27.5	0.90	−2.282	28.8	0.95	−1.830	29.8	0.96	−1.764	29.4	0.91	−2.206	28.4	0.93
7	−2.179	29.6	0.85	−2.063	31.1	0.90	−1.607	32.3	0.91	−1.565	31.7	0.86	−1.995	30.6	0.88
8	−2.002	31.4	0.81	−1.864	33.1	0.85	−1.402	34.5	0.88	−1.383	33.7	0.83	−1.803	32.7	0.84
9	−1.842	33.1	0.77	−1.683	35.0	0.82	−1.212	36.6	0.85	−1.215	35.7	0.80	−1.627	34.5	0.81
10	−1.694	34.6	0.74	−1.516	36.8	0.79	−1.033	38.6	0.83	−1.057	37.5	0.78	−1.464	36.2	0.78
11	−1.557	36.1	0.72	−1.361	38.4	0.76	−0.864	40.5	0.80	−0.907	39.2	0.76	−1.312	37.9	0.76
12	−1.429	37.4	0.69	−1.216	39.9	0.74	−0.701	42.3	0.79	−0.762	40.9	0.75	−1.169	39.4	0.74
13	−1.308	38.7	0.67	−1.080	41.3	0.71	−0.545	44.0	0.78	−0.623	42.5	0.74	−1.033	40.8	0.72
14	−1.193	39.9	0.66	−0.952	42.7	0.70	−0.394	45.7	0.76	−0.487	44.0	0.73	−0.904	42.2	0.70
15	−1.084	41.0	0.64	−0.830	44.0	0.68	−0.246	47.3	0.76	−0.353	45.6	0.72	−0.780	43.5	0.69
16	−0.980	42.1	0.63	−0.714	45.2	0.66	−0.101	48.9	0.75	−0.221	47.1	0.72	−0.661	44.8	0.67
17	−0.880	43.1	0.62	−0.604	46.3	0.65	0.042	50.5	0.75	−0.090	48.6	0.72	−0.546	46.0	0.66
18	−0.784	44.1	0.61	−0.497	47.5	0.64	0.184	52.0	0.75	0.041	50.1	0.72	−0.435	47.2	0.65
19	−0.691	45.1	0.60	−0.394	48.5	0.63	0.327	53.6	0.75	0.173	51.6	0.72	−0.326	48.3	0.65
20	−0.601	46.0	0.59	−0.295	49.6	0.62	0.470	55.2	0.75	0.306	53.1	0.72	−0.219	49.5	0.64
21	−0.513	47.0	0.58	−0.198	50.6	0.61	0.615	56.8	0.76	0.440	54.7	0.73	−0.115	50.6	0.64
22	−0.428	47.8	0.57	−0.104	51.6	0.61	0.762	58.4	0.77	0.577	56.3	0.74	−0.012	51.6	0.64
23	−0.344	48.7	0.57	−0.011	52.5	0.60	0.913	60.1	0.78	0.717	57.9	0.75	0.091	52.7	0.63
24	−0.263	49.6	0.56	0.081	53.5	0.60	1.069	61.8	0.79	0.861	59.5	0.76	0.193	53.8	0.63
25	−0.182	50.4	0.56	0.171	54.4	0.60	1.233	63.6	0.82	1.010	61.2	0.78	0.294	54.9	0.63
26	−0.103	51.2	0.55	0.261	55.4	0.59	1.405	65.5	0.84	1.168	63.0	0.80	0.397	56.0	0.64
27	−0.026	52.0	0.55	0.351	56.3	0.59	1.591	67.6	0.88	1.336	65.0	0.83	0.500	57.1	0.64
28	0.051	52.8	0.55	0.441	57.3	0.60	1.795	69.8	0.92	1.518	67.1	0.87	0.604	58.2	0.64
29	0.128	53.6	0.54	0.532	58.2	0.60	2.023	72.3	0.99	1.719	69.4	0.92	0.710	59.3	0.65
30	0.204	54.4	0.54	0.623	59.2	0.60	2.290	75.3	1.07	1.947	72.0	0.98	0.819	60.5	0.66
31	0.279	55.2	0.54	0.717	60.2	0.61	2.614	78.8	1.20	2.212	75.0	1.06	0.930	61.6	0.67
32	0.355	56.0	0.54	0.812	61.2	0.62	3.034	83.5	1.39	2.532	78.7	1.18	1.045	62.9	0.68
33	0.431	56.8	0.54	0.910	62.2	0.63	3.644	90.2	1.76	2.943	83.4	1.37	1.165	64.1	0.69
34	0.507	57.6	0.55	1.011	63.2	0.64	4.531	100.0	2.45	3.535	90.2	1.73	1.291	65.5	0.71
35	0.583	58.4	0.55	1.116	64.3	0.65	-	-	-	4.385	100.0	2.40	1.423	66.9	0.73
36	0.660	59.2	0.55	1.226	65.5	0.66	-	-	-	-	-	-	1.564	68.4	0.76
37	0.739	60.0	0.56	1.342	66.7	0.68	-	-	-	-	-	-	1.715	70.0	0.79
38	0.818	60.8	0.56	1.465	68.0	0.71	-	-	-	-	-	-	1.881	71.7	0.82
39	0.900	61.7	0.57	1.597	69.4	0.73	-	-	-	-	-	-	2.064	73.7	0.87
40	0.983	62.6	0.57	1.741	70.9	0.76	-	-	-	-	-	-	2.271	75.9	0.93
41	1.069	63.5	0.58	1.898	72.5	0.81	-	-	-	-	-	-	2.512	78.4	1.02
42	1.157	64.4	0.59	2.074	74.4	0.85	-	-	-	-	-	-	2.804	81.5	1.13
43	1.249	65.3	0.61	2.274	76.5	0.92	-	-	-	-	-	-	3.182	85.5	1.32
44	1.344	66.3	0.62	2.509	78.9	1.00	-	-	-	-	-	-	3.736	91.4	1.68
45	1.444	67.4	0.64	2.795	81.9	1.12	-	-	-	-	-	-	4.547	100.0	2.35
46	1.550	68.5	0.65	3.168	85.8	1.31	-	-	-	-	-	-	-	-	-
47	1.663	69.7	0.68	3.716	91.6	1.67	-	-	-	-	-	-	-	-	-
48	1.785	70.9	0.70	4.523	100.0	2.35	-	-	-	-	-	-	-	-	-
49	1.917	72.3	0.74	-	-	-	-	-	-	-	-	-	-	-	-
50	2.062	73.8	0.77	-	-	-	-	-	-	-	-	-	-	-	-
51	2.224	75.5	0.82	-	-	-	-	-	-	-	-	-	-	-	-
52	2.410	77.5	0.89	-	-	-	-	-	-	-	-	-	-	-	-
53	2.629	79.7	0.97	-	-	-	-	-	-	-	-	-	-	-	-
54	2.899	82.6	1.09	-	-	-	-	-	-	-	-	-	-	-	-

(Continued)

Table 3. (Continued).

Raw score	Motivation (BMQ-SF19)				Emotional Regulation (BREQ-SF17)				Social Cognition (BSCQ-SF13)				Disinhibition (BDQ-SF13)				Impulsivity (BIQ-SF16)			
	Logit scale	±95% CI	0–100 m-L scale	±95% CI	Logit scale	±95% CI	0–100 m-L scale	±95% CI	Logit scale	±95% CI	0–100 m-L scale	±95% CI	Logit scale	±95% CI	0–100 m-L scale	±95% CI	Logit scale	±95% CI	0–100 m-L scale	±95% CI
55	3.253	1.28	86.3	3.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
56	3.782	1.63	91.8	4.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
57	4.570	2.31	100.0	6.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Abbreviations: CI, confidence interval (equal to 1.96 standard error of measurement); m-L scale, modified logit scale

Notes:

The person estimates were reported; they are expressed both in logits and in a 0 to 100 modified-logit scale.

This conversion table can be used only if the patient responds to all items of the BIRT-PQs and if the modified questionnaires for BIRT-PQ-s are to be used, as detailed in Appendix 2.
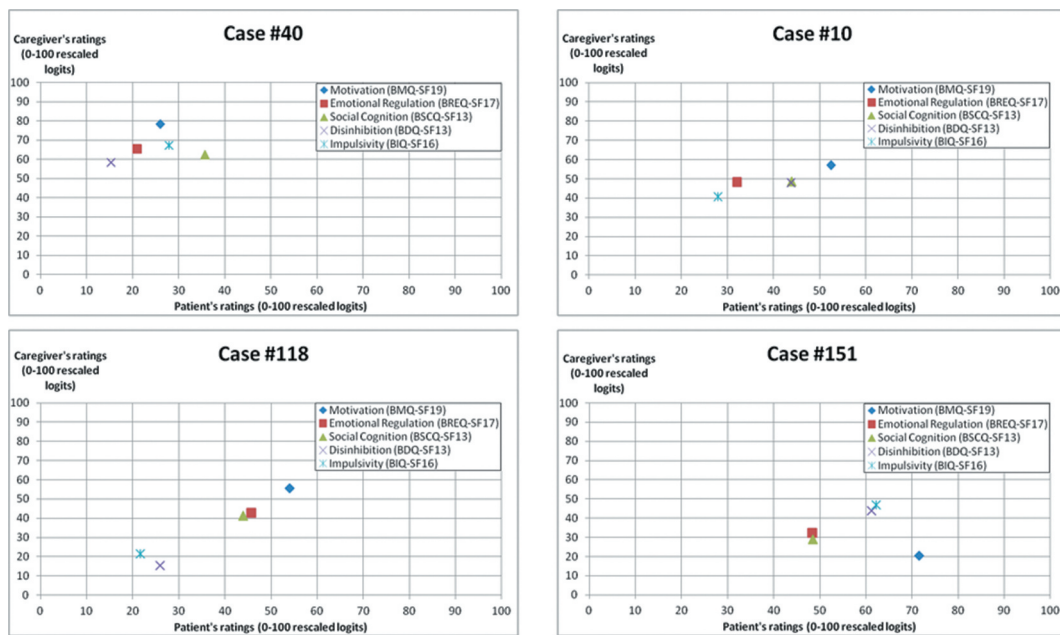
Figure 2. BIRT-PQ patient-caregiver profiles for four different cases in the sample. To construct these charts, the logit estimates of the five BIRT-BP were all rescaled into a 0–100 scale. For each BIRT-PQ, the patient's rating (horizontal axis) was plotted against the caregiver's rating (vertical axis). In this way, it is possible to appraise the degree of agreement (or disagreement) between the two ratings visually. For instance, the ratings for cases #10 and #118 are quite reasonably similar. On the other hand, for case #40 (top left corner), all caregiver ratings are markedly higher than the patient's rating. This implies the patient's underestimation and/or caregiver overestimation of the patient's neurobehavioral disability. For case #151, the opposite situation is evident: the patient's ratings are significantly higher than the caregiver's ones. This implies the patient's overestimation and/or caregiver's underestimation of the patient's neurobehavioral disturbances. Also, by these individual patients charts, it is possible to appraise the hierarchy of the neurobehavioural disability severity captured by the five BIRT-PQ. For instance, for Case #118, emotional disturbances are the second most prominent problem, whereas, for case #10, social cognition and disinhibition (overlapped) are the second most problematic issues. This knowledge may provide clinicians essential clinical information for personalizing the assessment and treatment strategy.

telephone' (BSCQ04), '*I get instructions wrong*' (BREQ05), '*I say things at the wrong time*' (BSCQ14), and '*I misunderstand people*' (BSCQ18), which were all linked to the 'inability to interpret external cues' facet. As BSCQ18 and BSCQ14 were also locally dependent with BSCQ25 ('*I say the wrong thing*'), only BSCQ04 was retained, because it fitted the model better than BSCQ05. BSCQ19 ('*I worry about what other people think* ') was deleted because severely misfitting the model, as it was an item which could be equally linked to the 'theory of mind/lack of empathy' and the 'social anxiety' facets, with a stronger emphasis than all the other items on the emotional components ('being worried').

After deleting 15 items (80% for LD), the remaining 13 items (BSCQ-SF13) satisfied the model's expectations in terms of stochastic invariance (χ^2_{52} = 47.1; p = .675), monotonicity, local independence, strict unidimensionality (PST = 5.0%; lower BCI = 3.3%), and absence of DIF for the tested group factors. All items fitted the model individually. At the content level, all four conceptual facets were represented. The item hierarchy showed that the facet associated with the lower levels of social cognition impairment was 'social anxiety' (average item difficulty: −0.587 logits). 'Difficulty in social interaction' (+0.091 logits) and 'theory of mind/lack of empathy' (+0.304 logits) were, on average, associated to increasing levels of difficulties with social cognition, whereas 'inability to interpret external cues' (average item difficulty in logits: +0.767) was the facet associated to the higher degree of

emotional dysregulation. However, there was only one item left linked to this facet.

The TI was −2.938, suggesting that the instrument was off-target for this sample, as also confirmed by the visual analysis of the targeting graph (Figure 1c). Although the PSI (0.783) was within the minimum cutoff for group measurement, the DI-PSI was 0.962, indicating the precision of measurement at the individual level (Table 2). The item fit statistics and the scoring model for the BSCQ-SF13 are reported in Appendix 2 c. The total score of the BSCQ-SF ranged from 0 to 34 points (Table 3).

BIRT disinhibition questionnaire (BDQ)

Within the BDQ, we identified three main conceptual facets: 'inhibition of behavior/delaying gratification,' 'inability to inhibit verbal behavior/lack of tact,' and 'sexual disinhibition.'

As reported in Table 2, the first Rasch analysis demonstrated that the data failed to meet the requirement of stochastic invariance (χ^2_{96} = 563.3; p = .000). The scale also failed the ICV requirements of monotonicity (ten items had DT), local independence (17 items had residual correlation values above the LDRC, here set at 0.165), unidimensionality (lower BCI = 5.2%). Furthermore, four items (BDQ21, BDQ11, BDQ12, and BDQ15) also failed the requirement of the absence of DIF. Finally, three items misfitted the model individually: BDQ11 and BDQ12 under-fitted the model with a fit residual

of 12.710 and 4.533, respectively, while BDQ10 over-fitted the model with a fit residual of −2.559.

After rescoring all items displaying DT, the item, '*I feel I have to do things even though I might get into trouble*' (BDQ11), was deleted immediately. This item was clearly misfitting (χ^2 = 158.7, fit residual >9) possibly because also influenced by the 'inability to foresee outcomes' facet, which is part of the impulsivity construct. Also, this item displayed a marked DIF by the respondent, as caregivers were more likely to affirm a higher score (indicative of a higher level of perceived disinhibition) than patients, given the same level of the construct. As for the previous questionnaires, other items were deleted only because of LD. For instance, considering '*It is hard to stop myself from doing things I know I should not do*' (BDQ03) and '*I do things that I know are wrong*' (BDQ04), which were both linked to the 'inhibition of behavior/delaying gratification' facet, only the former was retained because fitting the model better than the latter.

After deleting 11 items (seven for LD, four for misfit), the remaining 13-item scale (BDQ-SF13) satisfied all model's requirements in terms of stochastic invariance (χ^2_{52} = 64.5; p = .114), monotonicity, local independence, unidimensionality (PST = 4.4%; lower BCI = 2.7%), and absence of DIF (no item bias for all the tested subgroups). At the content level, all three facets were represented, although the items linked to the 'Inhibition of verbal behavior/lack of tact' facet were as twice as much of the items of the other two facets. The item hierarchy showed that the 'inhibition of verbal behavior/lack of tact' (average item difficulty: −0.270 logits) was the facet associated with the lower levels of disinhibition. On the other hand, 'inhibition of behavior/delaying gratification' (+0.177 logits) and 'sexual disinhibition' (+0.601 logits) were the facets associated with the higher levels of disinhibition. In particular, the three most difficult items were those related to socially-inappropriate behaviors, such as '*I get over-excited*' (BDQ06), '*I say rude things to people I do not know very well*' (BDQ05), and '*I hug and kiss strangers*' (BDQ14), where BDQ06 and BDQ14 were linked to the 'sexual disinhibition' facet.

The targeting graph demonstrated that the sample was not well-targeted to the sample (Figure 1d), as confirmed by a TI of −2.938. Although the PSI (0.747) was also adequate for group measurement in this case, the DI-PSI (0.962) was again indicating that the scale was precise enough for individual measurements (Table 2). The item fit statistics and the scoring model for the BDQ-SF13 are reported in Appendix 2d. The total score of BDQ-SF13 ranged from 0 to 35 points (Table 3).

BIRT impulsivity questionnaire (BIQ)

We identified four main conceptual facets within the BIQ: 'acting/speaking on impulse,' 'emotional impulsivity,' 'lack of planning/inability to foresee outcomes,' and 'snap decision-making/excessive spontaneity.' One item ('I find it hard to concentrate for a long time,' BIQ21) could not be linked univocally to any of the above facets.

As shown in Table 2, the initial Rasch analysis of the BIQ showed that the scale failed to meet most of the model's requirements in terms of stochastic invariance (χ^2_{96} = 609.1; $p \leq 0.001$), monotonicity (four items had DT), local independence (25 items had residual correlations above the LDRC, here set at 0.172), and unidimensionality (PST = 11.0%; lower BCI = 9.4%). However, the requirement of the absence of DIF was satisfied. At the item level, three items (BIQ13, BIQ26, BIQ29) over-fitted the model (fit residuals ranging from −3.534 to −2.649), while the other three items (BIQ01, BIQ18, BIQ20) under-fitted the model (fit residuals ranging from 2.874 to 12.185).

After rescoring all items displaying DT, we dealt with clusters of locally dependent items. The first two items which were deleted were '*I plan ahead*' (BIQ01) and '*I make a plan first before I start a task*' (BIQ20), which were locally dependent within the 'lack of planning/inability to foresee outcome' facet. But they also misfitted the model, as well as '*I find it hard to concentrate for a long time*' (BIQ21), possibly because all three may be influenced by an external variable (i.e., facets of the motivation construct). Instead, in another cluster related to compulsive shopping, both '*I buy more than I need*' (BIQ27) and '*If I see something, I like I buy it straight away*' (BIQ23) were deleted. However, within the same cluster, '*I buy things I do not need*' (BIQ12) and '*I spend all of my money as soon as I get it*' (BIQ12) were not locally dependent and were both retained because the latter focused more on the 'inability to foresee outcomes,' unlike the other three items where the central theme was 'acting on impulse.'

After deleting 16 items (81% because of LD), the final 16-item scale (BIQ-SF16) showed to fit the model adequately (χ^2_{64} = 85.8; p = .036), and there were no issues in terms of monotonicity, LD, acceptable unidimensionality (PST = 5.3%; lower BCI = 3.6%), and no DIF for the tested groups. All items fitted the model individually. At the content level, all four conceptual facets were represented. The item hierarchy showed that the 'snap decision-making/spontaneity' (average item difficulty in logits: −0.344), and 'acting/speaking on impulse' (−0.038 logits) were the facets indicating, on average, lower levels of impulsivity, whereas 'emotional impulsivity' (average item difficulty in logits: +0.092) and 'lack of planning/inability to foresee outcomes (+0.152 logits) were the facets associated with the higher levels of impulsivity. The mean sample ability was lower than the mean item difficulty, indicating that the scale was off-target, as also demonstrated by a TI of −3.315 (Figure 1e). Although the PSI of 0.824 indicated that the instrument could be used for group measurement, a DI-PSI of 0.973 suggested precision of the scale at the individual level (Table 2). The item fit statistics and the scoring model for the BIQ-SF16 are reported in Appendix 2e. The total score of BIQ-SF16 ranged from 0 to 45 points (Table 3).

Discussion

In this paper, we undertook an analysis of the ICV of the five BIRT-PQs within the RMT framework. After having demonstrated in another study that the original versions of each BIRT-PQs were not unidimensional (11), here we performed extensive structural modifications of each questionnaire, which led to reducing the item load from 150 to 78 items, which is an overall 48% item reduction. After these modifications, each short form fitted the Rasch model, with no DIF for age, gender, time since brain injury, etiology, and respondent. The

precision of measurement was sufficient for individual person measurements across all questionnaires, and total score to interval-level transformation tables for each scale were made available (Table 3). In this way, the correct computation of change scores and the use of parametric statistics are now allowed. The 48% reduction of the item load yielded a significant decrease in the administration time for both patients and caregiver versions.

Given the widespread violations of the local independence and unidimensionality requirements revealed by the CFA (11), the base Rasch analyses showed a lack of fitness to the Rasch model for all the original scales. However, the base Rasch analyses also highlighted widespread violations of other relevant measurement requirements for all the scales, i.e., violations of monotonicity and presence of DIF, along with several misfitting items. As one goal of this study was to reduce the item load, the analysis strategy focused, in the first instance, on structural modifications (i.e., item rescoring for violations of monotonicity and item deleting for misfitting items or other ICV violations).

About 22% of the items within the whole item set required rescoring, due to DT. However, as having different scoring patterns within a questionnaire may be a nuisance from a clinical point of view, an earlier (not reported) analytical strategy aimed at maintaining the original scoring format across all items. Thus, within this strategy, we did not perform any item rescoring, but we just deleted items displaying violations one or more of the ICV requirements. However, after dealing with LD, there were still items with disordered thresholds. Despite rescoring or removing these items, it was not possible to reach a solution for any of the BIRT-PQ. Therefore, we reported the current analyses where the item rescoring was performed first, followed by the item deletion. We believe that thresholds reversal observed for some of the items were caused by the low endorsement frequencies in the middle scoring categories due to the skewed distributions of the samples (54). However, it is highly likely that also item redundancy played a role, as LD itself may lead itself to the disordering of the thresholds (55).

Mainly, the item deletion procedure was initially focused on accounting for LD by eliminating one item in the locally dependent pair. Within the RMT, LD may arise because of trait dependency (which entails multidimensionality, as the items' responses are both influenced by an external variable) or response dependency (where the responses to an item in the pair are dependent on the responses to the other item). In both cases, we addressed LD extensively, as it severely distorts the measurement properties of the scale. After accounting for all LD, we addressed with item deletion any further violations of the model's requirements (i.e., the presence of DIF and item misfit).

Within this paper, we linked the items of the BIRT-PQ to the original concepts suggested by Hyde (30), given the lack of published data on the precise content validity of the questionnaires. As this procedure uncovered several unsolvable linking inconsistencies and difficulties, we decided to collapse conceptual categories deemed to be too similar to devise the minimum conceptual category set represented by each BIRT-PQ. Only in some cases, the obtained theoretical set resembled that of established neuropsychological models (e.g., the apathy

model by Levy and Dubois (56) for the motivation questionnaire). Given the exquisite confirmatory nature of the Rasch model, the conceptual analysis performed was a valuable guiding tool for the item reduction procedure and to interpret the results. In particular, it allowed us: a) to uncover that most of the LD occurred within the same conceptual facets, thus providing substantive evidence of the hypothesized redundancy of the questionnaires; b) to appraise the multidimensional nature of some items easily; c) to interpret the item hierarchy suggested by the analysis in conceptual terms, which may be particularly useful for clinicians.

Following these extensive structural modifications, the final versions of the reduced BIRT-PQs fitted the Rasch model adequately. Furthermore, these shorter forms are free from DIF and, therefore, can be administered regardless of the patient's age, gender, etiology, time since the brain injury, and typology of the respondent (patient vs. caregiver). Additionally, the fact that all items are free from DIF by respondent implies that any substantial measurement difference between the estimates obtained within a pair of matched respondents (a patient and the corresponding caregiver) will reflect real measurement differences in the latent variable rather than item bias, as was shown in Figure 2. The profile of the patient-caregiver BIRT-PQ estimates demonstrates the clinical utility of the measure, as it may provide clinicians with essential clinical information for personalizing the treatment strategy.

The reliability of the final solutions for BSCQ, BDQ, and BIQ scales revealed that some of these shorter forms did not reach the minimum cutoff for individual person measurements (i.e., 0.850). This low reliability may be the consequence of the observed skewed distribution of the persons affirming the questionnaires, as this is known to affect the PSI values. For this reason, we estimated a separation reliability index (DI-PSI), which is independent of the distribution of the sample and which provided a more conservative value of reliability. The DI-PSI suggested that the precision of each scale was sufficient (>0.850) for individual-patient clinical decision-making purposes.

The deletion of several items of the original version is likely to facilitate the administration of the instruments. Indeed, we estimate that a 48% reduction of the item load could yield a reduction of the administration time from 32 to 15.4 minutes, and from 23.8 to 11.4 minutes for the patient and caregiver versions, respectively. This substantial reduction of the administration time could significantly improve the feasibility and acceptability of these instruments (57), thus facilitating their use in daily clinical practice. Furthermore, we highly recommend that clinicians and researchers make use not only of the shorter versions of the BIRT-PQs (provided in Appendix 3 and 4) but also of the raw score to interval-level measure transformation tables provided in Table 3. These interval measures fully support the use of parametric statistics required by clinical trials and can be used to measure the change accurately over time (15,28,58).

This study has some limitations. First, although the sample size was sufficient for stable calibrations of the final shorter BIRT-PQs versions, it was too small to have a 'set aside' sample, which would have enabled us to validate the final questionnaires further. Thus, there is a risk that the solutions obtained have

capitalized on chance regarding fit to the model. Consequently, these findings require replication. Second, the conceptual content analysis of the BIRT-PQ was performed with the only aim of facilitating the interpretation of the results from a clinical point of view. As we used a modified version of the original content categories for the BIRT-PQ, there may be some conceptual imprecisions and inaccuracies when compared with established neuropsychological models of the constructs involved. Thus, the conceptual hierarchy suggested by the item ordering for each final solution should be interpreted cautiously. Third, as we enrolled only adult participants with s-ABI, the findings of this study apply only to this population. Fourth, as only Italian speaking participants with s-ABI and their respective caregivers were recruited, the generalizability of the results to people living in other countries may be limited. Further international studies enrolling people living in different countries may be useful to assess if the questionnaires operate in the same manner across different cultures and languages.

Conclusion

We demonstrated that the total scores of the five original BIRT-PQs were invalid in terms of ICV, both from the perspective of CTT (11) and RMT, given severe violations of the local independence, unidimensionality, and invariance requirements. In this study, after extensive structural modifications, including a significant reduction of the item load, we obtained five short-forms of the BIRT-PQs satisfying the strict measurement requirements of the RMT. These SF can be used in clinical practice and research to measure several dimensions of the perception of the neurobehavioral disability in adults with s-ABI and their caregivers. Further studies, making use of the provided interval score transformations, are needed to investigate the external construct validity of these new measures, to establish normality cutoffs, and to investigate the clinical advantages and significance of the availability of two different estimates (patient and caregiver) on the same metric of a patient's neurobehavioral disability.

Data availability statement

The data that support the findings of this study may be downloaded from https://www.dropbox.com/sh/dzq85x14yovrlej/AAD4HHb3meoJue7Nyk4sgng_a?dl=0.

Disclosure statement

The authors report no conflict of interest

ORCID

Leonardo Pelicciari http://orcid.org/0000-0002-9521-3759
Daniele Piscitelli http://orcid.org/0000-0002-5240-3798
Benedetta Basagni http://orcid.org/0000-0002-6173-584X
Antonio De Tanti http://orcid.org/0000-0003-4803-419X
Serena Caselli http://orcid.org/0000-0002-4187-9475
Maria Paola Ciurli http://orcid.org/0000-0002-2019-9569
Anna Estraneo http://orcid.org/0000-0001-6646-5626
Maria Grazia Gambini http://orcid.org/0000-0002-6211-2169
Gianfranco Lamberti http://orcid.org/0000-0002-3288-5275
Mauro Mancuso http://orcid.org/0000-0002-8451-9277
Matteo Sozzi http://orcid.org/0000-0002-3969-7330
Laura Abbruzzese http://orcid.org/0000-0002-5304-3974
Fabio La Porta http://orcid.org/0000-0002-5031-6658

References

1. Wood RL. Understanding neurobehavioural disability. In: Wood RL, McMillan TM, editors. Neurobehavioural disability and social handicap following traumatic brain injury. Philadelphia: Psychology Press Ltd; 2001. p. 3–27.
2. Cattran CJ, Oddy M, Wood RL, Moir JF. Post-injury personality in the prediction of outcome following severe acquired brain injury. Brain Inj. 2011;25(11):1035–46. doi:10.3109/02699052.2011.607787.
3. Nightingale EJ, Soo CA, Tate RL. A systematic review of early prognostic factors for return to work after traumatic brain injury. Brain Impair. 2007;8(2). doi:10.1375/brim.8.2.101.
4. Moretta P, Masotta O, Crispino E, Castronovo G, Ruvolo S, Montalbano C, Loreto V, Trojano L, Estraneo APsychological distress is associated with altered cognitive functioning in family caregivers of patients with disorders of consciousness. Brain Inj. 2017;31(8):1088–93. doi:10.1080/02699052.2017.1290278.
5. Moretta P, Estraneo A, De Lucia L, Cardinale V, Loreto V, Trojano L.A study of the psychological distress in family caregivers of patients with prolonged disorders of consciousness during in-hospital rehabilitation. Clin Rehabil. 2014;28(7):717–25. doi:10.1177/0269215514521826.
6. Kreutzer JS, Gervasio AH, Camplair PS. Primary caregivers' psychological status and family functioning after traumatic brain injury. Brain Inj. 1994;8(3):197–210. doi:10.3109/02699059409150973.
7. Kaufer DI. Neurobehavioral assessment. Continuum (Minneap Minn). 2015;21(3Behavioral Neurology and Neuropsychiatry):597–612. doi:10.1212/01.CON.0000466655.51790.2f.
8. Silver JM, McAllister TW, Arciniegas DB. American psychiatric association. Textbook of traumatic brain injury. 3rd ed. Washington (DC): American Psychiatric Association Publishing; 2019. p. xxxii, 953.
9. Oddy M, Cattran C, Wood R. The development of a measure of motivational changes following acquired brain injury. J Clin Exp Neuropsychol. 2008;30(5):568–75. doi:10.1080/13803390701555598.
10. Cattran C, Oddy M, Wood R. The development of a measure of emotional regulation following acquired brain injury. J Clin Exp Neuropsychol. 2011;33(6):672–79. doi:10.1080/13803395.2010.550603.
11. Basagni B, Piscitelli D, De Tanti A, Pellicciari L, Algeri L, Caselli S The unidimensionality of the five brain injury rehabilitation trust personality questionnaires (BIRT-PQs) may be improved: preliminary evidence from classical psychometrics. Brain Inj. 2020:1–12. doi:10.1080/02699052.2020.1723700.
12. Kucukdeveci AA, Tennant A, Grimby G, Franchignoni F. Strategies for assessment and outcome measurement in physical and rehabilitation medicine: an educational review. J Rehabil Med. 2011;43(8):661–72. doi:10.2340/16501977-0844.
13. La Porta F, Franceschini M, Caselli S, Cavallini P, Susassi S, Tennant A. Unified Balance Scale: an activity-based, bed to community, and aetiology-independent measure of balance calibrated

with Rasch analysis. J Rehabil Med. 2011;43(5):435–44. doi:10.2340/16501977-0797.

14. Holland PW, Wainer H. Differential item functioning, Holland PW, Wainer H, editors. Hillsdale (NJ): Lawrence Erlbaum Associates, Inc; 1993. xv, 453p.

15. 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(8):1358–62. doi:10.1002/art.23108.

16. 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(Pt 1):1–18.

17. Andrich D. Rasch models for measurement. Newbury Park: Sage Publications; 1988. p. 95.

18. Rasch G. Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. 1960.

19. Kreiner S. The Rasch models for dichotomous items. London, Hoboken (NJ): ISTE; John Wiley & Sons; 2013. p. xvi, 368.

20. 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(2):205–31.

21. Marais I. Local dependence. London Hoboken (NJ): ISTE; John Wiley & Sons; 2013. p. xvi, 368.

22. Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, Lawton G, Simone A, Carter J, Lundgren-Nilsson Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care. 2004;42(1 Suppl):I37–48. doi:10.1097/01.mlr.0000103529.63132.77.

23. Fisher WP. Rating scale instrument quality criteria. Rasch Meas Trans. 2007;21(1):1095.

24. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess. 2009;13(12):iii,ix–x, 1–177. doi:10.3310/hta13120.

25. Embretson SE, Reise SP. Item response theory for psychologists. Mahway (NJ): Lawrence Erlbaum Associates; 2000.

26. Perline R, Wright BD, Wainer H. The Rasch model as additive conjoint measurement. Appl Psychol Meas. 1979;3(2):237–55.

27. Tesio L. Measuring behaviours and perceptions: rasch analysis as a tool for rehabilitation research. J Rehabil Med. 2003;35 (3):105–15.

28. Piscitelli D, Pellicciari L. Responsiveness: is it time to move beyond ordinal scores and approach interval measurements? Clin Rehabil. 2018;32(10):1426–27. doi:10.1177/0269215518794069.

29. World Medical Association. Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310(20):2191–94. doi:10.1001/jama.2013.281053.

30. Hyde CJ. The measurement and impact of five personality changes after brain injury. Swansea: University of Wales; 2006.

31. Basagni B, Navarrete E, Bertoni D, Cattran C, Mapelli D, Oddy M, De Tanti A. The Italian version of the brain injury rehabilitation trust (BIRT) personality questionnaires: five new measures of personality change after acquired brain injury. Neurol Sci. 2015;36 (10):1793–98. doi:10.1007/s10072-015-2251-9.

32. Pellicciari L, Piscitelli D, Caselli S, La Porta FA. Rasch analysis of the Conley Scale in patients admitted to a general hospital. Disabil Rehabil. 2018;1–10. doi:10.1080/09638288.2018.1478000.

33. 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(7):1209–16. doi:10.1016/j.apmr.2012.02.020.

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

(6):705–16.

35. Panella L, La Porta F, Caselli S, Marchisio S, Tennant A. Predicting the need for institutional care shortly after admission to rehabilitation: rasch analysis and predictive validity of the BRASS Index. Eur J Phys Rehabil Med. 2012;48(3):443–54.

36. Andrich D, Sheridan BS, Luo G. RUMM 2030: Rasch unidimensional models for measurement manual (version 5.1). Perth, Western Australia: RUMM Laboratory; 2010.

37. Maritz R, Tennant A, Fellinghauer C, Stucki G, Prodinger 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(3):193–200. doi:10.2340/16501977-2525.

38. Pellicciari L, Ottonello M, Giordano A, Albensi C, Franchignoni F. The 88-item multiple sclerosis spasticity scale: a Rasch validation of the Italian version and suggestions for refinement of the original scale. Qual Life Res. 2019;28 (1):221–31. doi:10.1007/s11136-018-2005-2.

39. Meroni R, Piscitelli D, Bonetti F, Zambaldi M, Cerri CG, Guccione AA, Pillastrini P. Rasch analysis of the italian version of pain catastrophizing scale (PCS-I). J Back Musculoskelet Rehabil. 2015;28(4):661–73. doi:10.3233/BMR-140564.

40. Christensen KB, Makransky G, Horton M.Critical values for Yen's Q 3: identification of local dependence in the Rasch model using residual correlations. Appl Psychol Meas. 2017;41(3):178–94. doi:10.1177/0146621616677520.

41. Tennant A, Pallant J. DIF matters: A practical approach to test if differential item functioning makes a difference. Rasch Measure Trans. 2007;20(4):1082–84.

42. Wright BD. Separation, reliability and skewed distributions: statistically different levels of performance. Rasch Measure Trans. 2001;14(4):786.

43. Brodersen J, Doward LC, Thorsen H, Mckenna SP. Writing health-related items for Rasch models - patient-reported outcome scales for health sciences: from medical paternalism to patient autonomy. In: Christensen KB, Kreiner S, Mesbah M, editors. Rasch models in health. applied mathematics series, 281–302. London UK, Hoboken NJ: ISTE Ltd and John Wiley & Sons, Inc; 2013.

44. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement information system (PROMIS). Med Care. 2007;45(5 Suppl 1):S22–31. doi:10.1097/01.mlr.0000250483.85507.04.

45. Revicki DA, Chen W, Tucker CA. Developing item banks for patient-reported health outcomes. In: Handbook of item response theory modeling: applications to typical performance assessments, 334–363. New York (NY): Routledge; 2014.

46. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas. 2002;3(1):85–106.

47. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. J Rehabil Med. 2010;42 (4):323–31. doi:10.2340/16501977-0537.

48. Geri T, Piscitelli D, Meroni R, Bonetti F, Giovannico G, Traversi R, Testa M. Rasch analysis of the neck bournemouth questionnaire to measure disability related to chronic neck pain. J Rehabil Med. 2015;47(9):836–43. doi:10.2340/16501977-2001.

49. Lundgren Nilsson A, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM) revisited. J Rehabil Med. 2011;43(10):884–91. doi:10.2340/16501977-0871.

50. Andrich D, Lyne A, Sheridan B, Luo G. RUMM 2020. Perth: RUMM Laboratory; 2003.

51. Linacre JM. Sample size and item calibration stability. Rasch Meas Trans. 1994;7:328.

52. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ. 1995;310(6973):170. doi:10.1136/bmj.310.6973.170.

53. La Porta F. on behalf of the ERRTG (European Rasch Research & Teaching Group). Bologna, Italy: RUMM Logbook v1.9.5 ed; 2018.

54. Andrich D. An expanded derivation of the threshold structure of the polytomous rasch model that dispels any "threshold disorder controversy". Educ Psychol Meas. 2013;73 (1):78–124. doi:10.1177/0013164412450877.

55. Andrich D, Humphry SM, Marais I. Quantifying local, response dependence between two polytomous items using the Rasch model. Appl Psychol Meas. 2012;36(4):309–24. doi:10.1177/0146621612441858.

56. Levy R, Dubois B. Apathy and the functional anatomy of the prefrontal cortex-basal ganglia circuits. Cereb Cortex. 2006;16 (7):916–28. doi:10.1093/cercor/bhj043.

57. Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, Williamson PR, Terwee CB. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. Trials. 2016;17(1):449. doi:10.1186/s13063-016-1555-2.

58. Svensson E. Guidelines to statistical evaluation of data from rating scales and questionnaires. J Rehabil Med. 2001; 33(1650–1977(Print)):47–48. doi:10.1080/165019701300006542.