ORIGINAL REPORT

Check for updates

DIMENSIONALITY AND SCALING PROPERTIES OF THE PATIENT CATEGORISATION TOOL IN PATIENTS WITH COMPLEX REHABILITATION NEEDS FOLLOWING ACQUIRED BRAIN INJURY

Richard J. SIEGERT, PhD[1], Oleg MEDVEDEV, PhD[2] and Lynne TURNER-STOKES, DM, FRCP[3,4]

From the [1]Auckland University of Technology, School of Public Health & Psychosocial Studies and School of Clinical Sciences, [2]University of Auckland, School of Medicine, Centre for Medical and Health Sciences Education, New Zealand, [3]Northwick Park Hospital, London North West Healthcare NHS Trust, and [4]Department of Palliative Care Policy and Rehabilitation, School of Nursing Midwifery and Palliative Care, King's College London, London, UK

Objective: To investigate the scaling properties of the Patient Categorisation Tool (PCAT) as an instrument to measure complexity of rehabilitation needs.
Design: Psychometric analysis in a multicentre cohort from the UK national clinical database.
Patients: A total of 8,222 patents admitted for specialist inpatient rehabilitation following acquired brain injury.
Methods: Dimensionality was explored using principal components analysis with Varimax rotation, followed by Rasch analysis on a random sample of $n = 500$.
Results: Principal components analysis identified 3 components explaining 50% of variance. The partial credit Rasch model was applied for the 17-item PCAT scale using a "super-items" methodology based on the principal components analysis results. Two out of 5 initially created super-items displayed signs of local dependency, which significantly affected the estimates. They were combined into a single super-item resulting in satisfactory model fit and unidimensionality. Differential item functioning (DIF) of 2 super-items was addressed by splitting between age groups (< 65 and ≥ 65 years) to produce the best model fit ($\chi^2/df = 54.72$, $p = 0.235$) and reliability (Person Separation Index (PSI) = 0.79). Ordinal-to-interval conversion tables were produced.
Conclusion: The PCAT has satisfied expectations of the unidimensional Rasch model in the current sample after minor modifications, and demonstrated acceptable reliability for individual assessment of rehabilitation complexity.

Acquired brain injury (ABI) typically results in a diverse range of physical, cognitive, and psychosocial impairments, and patients may have widely differing needs for rehabilitation. In the UK, NHS England identifies 3 levels of inpatient rehabilitation service (1–3) and 4 categories of patient need (A–D) (1). The majority of patients have category C or D needs, which can be met by their local general (Level 3) rehabilitation services. Patients with more complex (category B) needs may require treatment in a local specialist (Level 2) rehabilitation unit; and a small number will have highly complex (category A) needs requiring the specialist skills and facilities of a tertiary (Level 1) rehabilitation service. The UK Rehabilitation Outcomes Collaborative (UKROC) Database provides the national clinical dataset for commissioning of specialist inpatient rehabilitation services and reporting of the dataset is now mandatory for all Level 1 and 2 rehabilitation services in England.

"Case complexity" in rehabilitation may have several meanings. At one level it may reflect the need for resources (in terms of staff inputs or cost) or it may reflect the wider aspects of the biopsychosocial model (2). Within the UKROC database the Rehabilitation Complexity Scale (3) is designed to record rehabilitation resource requirements and to identify the costs of rehabilitation. However, it does not give any indication of what the staff may be spending their time on. The Patient Categorisation Tool (PCAT) is a more wide-ranging tool for identifying patients with complex (category A and B) needs requiring rehabilitation in a Level 1 or 2 service. It may be used to generate a "complexity profile" for a given patient, describing their different types of requirements, and this information may, in turn, help direct them to the most appropriate service to meet their needs. It now forms part of the UKROC dataset (4). Originally produced as a checklist it was subsequently developed as an ordinal scale. In addition to identifying those individuals who are likely to require specialist rehabilitation, it is also used at a population level as a numerical measure to rate and compare the complexity of the clinical caseload across different services. This forms part of the national bench-marking process that signposts rehabilitation units for designation at different levels of specialization (5). It is therefore important to understand its scaling and measurement properties in the various different patient groups presenting for rehabilitation.

Initial evaluation using traditional psychometric techniques in a general, but mainly neurological,

rehabilitation sample has shown the PCAT to have acceptable measurement properties with a single overall scale, subdivided into 2 main sub-scales, respectively, identifying "physical" and "cognitive/psychosocial" needs (6). A total PCAT score of ≥30 proved to be a reasonably sensitive and specific indicator of category A patients. However, there are residual concerns about summing ordinal data, especially across more than one domain.

Rasch analysis uses a probabilistic statistical model to evaluate the legitimacy of summing item scores to generate a single measurement. Proponents of the Rasch model argue that it provides a more rigorous assessment of scaling properties (7) than traditional methods, and it has been used to transform ordinal questionnaire scores to yield a genuine interval scale (8). The value of this approach has been demonstrated at the group level for such scales where scores are summed and compared across different subgroups (9–11). For example, in the context of rehabilitation, Rasch interval metrics have been used to transform total scores between different functional scales, such as the FIM™ Motor Scale and the Barthel Index, which may be valuable for meta-analysis and institutional benchmarking (12).

The Rasch model is based on the underpinning principle that performance on a scale is determined by just 2 parameters; the ability of the individual and the difficulty of the items. Rasch analysis examines the fit of the data to the predicted model. The logic of "item difficulty" and "individual ability" is immediately self-evident within the context of measurement of functional ability. However, it can also be translated to different constructs. The construct of measurement for the PCAT is "rehabilitation complexity", so in this case the Rasch model would be represented by the rehabilitation needs of the individual (person) and the resource requirements to meet them (item).

The purpose of this study was to examine the scaling properties of the PCAT tool in a large multi-centre cohort of patients with complex rehabilitation needs following acquired brain injury. Rasch analysis was used to determine whether the PCAT can be used as a unidimensional measure of complexity of rehabilitation needs in this population and to produce a transformation table for converting ordinal to interval data.

METHODS

Measure

The PCAT tool is designed to capture complex needs for rehabilitation in terms of the skills, time and facilities that may be found in a specialist rehabilitation setting. It comprises 18 items, each rated on a score of 1–3. When calculating the total score, the UKROC dataset takes only the higher of the scores for Medical and Neuropsychiatric inputs, and re-scores 1 item (Duration) on a scale of 0–2, to give a 17-item scale with a total score range of 16–50. The PCAT and some example scoring profiles are given in Appendix SI[1].

Data source and sampling

The PCAT was first introduced in 2012, but reporting was optional until 2015. Therefore complete data were not expected. The analysis was conducted on all episodes with a PCAT score recorded on admission in the UKROC database for patients admitted for rehabilitation in Level 1 and 2 services (*n*=67) between April 2012 and December 2016. Episodes with length of stay 8–400 days were selected to exclude patients admitted for assessment only or for long-stay programmes of care. Fig. 1 summarizes the process of extraction and analysis.

During this period, PCAT ratings were recorded for a total of 8,222 patients with acquired brain injury, comprising 56.3% of the total cohort. Within this dataset there were no missing item scores. A sub-sample of 500 episodes (sub-sample B) was
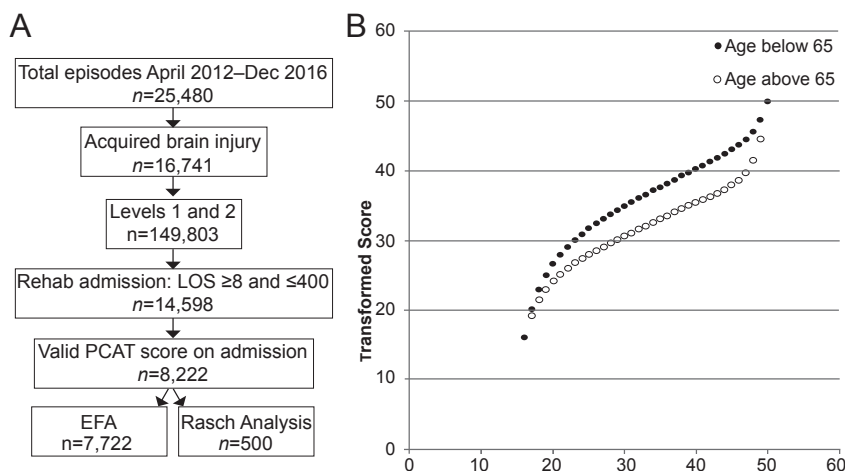
---

[1]https://doi.org/10.2340/16501977-2327

Fig. 1. Data extraction process to derive the dataset used for analysis. (EFA: exploratory factor analysis). (A) Flow chart of extraction. (B) Scatterplot of ordinal vs transformed scores. LOS: length of stay; PCAT: Patient Categorisation Tool.

randomly extracted for Rasch analysis, using the randomization facility in SPSS. Exploratory factor analysis was conducted on the remaining $n=7,722$ episodes (sub-sample A).

Ethics approval

The UKROC programme was registered as a multicentre service evaluation until July 2015 and is now commissioned directly by NHS England. Collection and reporting of the UKROC dataset is a commissioning requirement according to the NHSE service specification for Level 1 and 2 Rehabilitation Services. According to the UK Health Research Authority, the publication of research findings from de-identified data gathered in the course of routine clinical practice does not require research ethics permission.

Statistical analysis and software

Descriptive analysis was carried out using IBM SPSS v.23 software. Exploratory factor analysis was conducted to explore dimensionality. This was followed by Rasch analysis using RUMM2030 software (30) to test dimensionality. A significance value of 0.05 was used throughout.

For the exploratory factor analysis we applied principal components analysis (PCA) with Varimax rotation as these typically provide clear, interpretable solutions and have been used in our previous evaluations of the UKROC measurement tools (3, 13–15). The Kaiser Myer Olkin test and Bartlett's Test of Sphericity were used to make sure that the correlation matrix was suitable for factor analysis. The decision on the number of factors to rotate was based upon consideration of: (*i*) the number of factors with eigenvalues >1, (*ii*) visual inspection of the scree plot, and (*iii*) parallel analysis (16). Eigenvalues are numbers representing how much variance an individual component explains. (17) In the Scree test these are plotted against their components in descending order to help determine the fewest number of components explaining most of the variance (18). Parallel analysis involves determining the number of factors to rotate by comparing the size of the eigenvalues obtained with those from a Monte Carlo simulation (19). Internal consistency was examined using Cronbach's alpha.

An extensive literature provides guidance on methodology for Rasch analysis (8, 9, 20–22). Lundgren-Nilsson & Tennant 2011 (11) reviewed approaches to Rasch in the context of rehabilitation and made the following recommendations to improve the rigor of future analyses:

- Sample size should be a minimum 20 cases per item in the largest subscale or 243 participants, whichever value is larger (16, 23).
- Use of the Andrich Rating Scale vs the Partial Credit Model chosen according to the Likelihood-Ratio test. The Rating Scale model assumes the distances between thresholds to be the same across all items, whereas the Partial Credit allows for flexibility. A significant likelihood ratio test indicates the data are suitable for the Partial Credit Model.
- Examine several "analytical pathways", including some with and without re-ordering disordered thresholds.
- Creation of testlets or "super items" to deal with local response dependency. Local dependency is identified by correlations in the residual (unexplained) variance among items. Locally dependent items are simply summed or combined to form a super-item.

- Unidimensionality is tested using Rasch principal components analysis of the residuals and the equating test with paired *t*-tests across all participants.
- Where present Differential Item Functioning (DIF) might require splitting the sample according to the relevant person factor (e.g. age, sex, diagnostic group, etc.).
- Item removal only as a last resort (in order to maintain the clinical integrity of the instrument).
- Where possible, production of a transformation table to convert raw scores to Rasch-transformed scores, thus encouraging clinicians to use interval scores.

We followed all the above steps to deal with each of these issues, when they arose. The sample size of 500, representing approximately 30 cases per item, was chosen to ensure at least 100 cases in each of the diagnostic sub-groups (trauma, stroke and other). We used the likelihood ratio test, to determine whether the Rating Scale or partial credit model for Rasch analysis was most appropriate. The summary statistics of the Rasch model were assessed based on the mean item and person location, individual item fit residual, the overall item-trait interaction χ^2 test/*p*-value and the Person Separation Index (PSI), interpreted as follows:

- The mean item location (in this case reflecting resource requirements) is always set to zero.
- A mean person location (reflecting needs) of ± 0.5 indicates a well-targeted scale.
- The overall mean values representing perfect fit for both item and person fit residual are 0.0 (standard deviation (SD)=1).
- The item-trait interaction χ^2 reflects the overall fit of the data to the model's expectations and should be not significant ($p>0.05$).
- The PSI is a measure of internal consistency of the scale, similar to a Cronbach's alpha in classical test theory (30); PSI values above 0.7 are required for group use and above 0.8 for individual assessment.

We tested for item bias across important person factors, such as age group (<44, 45–64, 65 years plus), sex, and diagnostic category (i.e. trauma or stroke). Items displaying DIF were "split" to allow variation by the corresponding factor. This DIF split effectively completes a separate Rasch analysis for each person variable, where DIF is evident by splitting the sample accordingly (e.g. males and females separately). As it was desirable to keep the original structure of the PCAT scale, item removal would only be considered as a last resort to improve the fit. The items at risk of deletion were those exhibiting significant misfit, i.e. excessive residual values ($>\pm 2.5$) and a *p*-value significant at the 0.05 level, with a Bonferroni adjustment for multiple tests (24). Unidimensionality was tested using Rasch PCA of the residuals and the equating *t*-test. The PCA of the residuals examines for evidence of multidimensionality after removal of the "Rasch factor". The equating *t*-test creates 2 subtests from the first principal component and tests for any significant difference in person estimates between these subtests. Unidimensionality of the scale is confirmed if significant *t*-test comparisons do not exceed 5% or if the lower bound of a binominal confidence interval computed for the number of significant *t*-tests overlaps the 5% cut-off point (25). Finally, we followed the 10 quality indicators for evaluating the quality of reporting Rasch studies developed by the Rasch Special Interest Group of Outcome Measures in Rheumatology (OMERACT 11) (26).

Table I. Demographics of the total sample and randomized subsamples

	Total sample $n=8,222$	Sub-sample A $n=7,722$	Sub-sample B $n=500$
Age, years*	53.1 (±0.4)	53.1 (±0.4)	53.3 (±0.4)
Males:Females, %	61:39	61:39	63.37
Length of stay, days*	90 (±1.4)	89 (±1.4)	90 (±1.4)
Aetiology, *n* (%)			
Stroke	3,974 (48)	3,735 (48)	239 (48)
Trauma	2,184 (27)	2,042 (28)	142 (26)
Other	1,964 (22)	1,852 (22)	112 (24)
Service level, *n* (%)			
Level 1 ($n=18$)	3,012 (37)	2,838 (37)	174 (35)
Level 2 ($n=49$)	5,210 (63)	4,884 (63)	326 (65)
Total PCAT score*	31.6 (±0.1)	31.6 (±0.1)	31.0 (±0.1)
Median IQR	31 (27–36)	31 (27–36)	31 (27–35)

*mean (±95% confidence intervals (CI) are based on 1,000 bootstrapped samples.

IQR: interquartile range; PCAT: Patient Categorisation Tool.

RESULTS

The demographics of all 3 samples are shown in Table I. No significant differences were found between the 2 sub-samples.

Within this sample, 4,553 (55%) of patients were assessed as having category A needs, 3,367 (41%) as category B and just 295 (3.6%) as category C or D needs. The median total PCAT score for patients with category A needs was 35 (IQR 31–39); for category B needs 28 (25–30) and for category C/D 22 (19–25) (Fig. 2). Significant differences were seen between all groups (Mann–Whitney $p<0.001$). When total PCAT scores were analysed within the clinically-assessed categories of need, a total PCAT score ≥30 identified patients with category A needs with 84% sensitivity and 74% specificity.
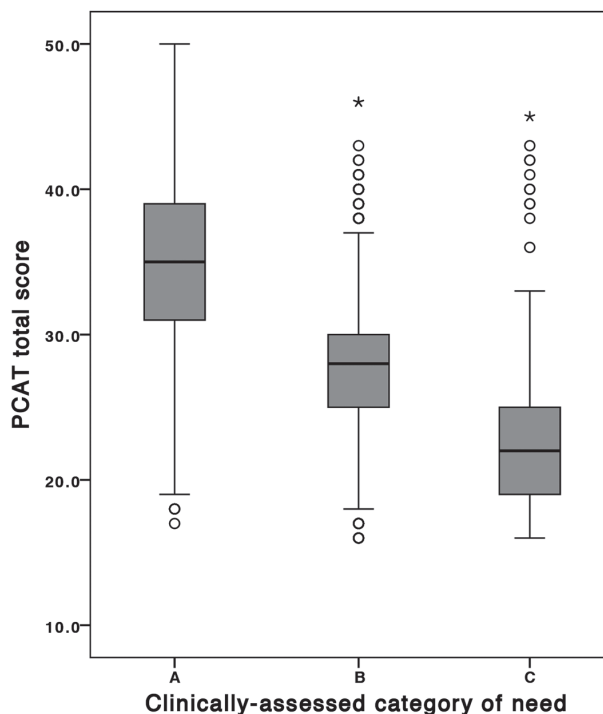

Fig. 2. Box plots of raw total scores between clinically-assessed category of need illustrate clear distinction between different clinical categories of rehabilitation need.

Principal components analysis: dimensionality and internal consistency

The results of exploratory factor analysis (sub-sample A) are summarized in Table II. Corrected item total correlations for the full scale were all significant at $p<0.001$ and ranged from 0.17 to 0.62 and Cronbach's

Table II. Results of principal component analysis

Item	Median (IQR) score	Item total correlation full scale**	Loading on 1st principal component	Varimax rotation Component 1	Component 2	Component 3	Corresponding super-item for Rasch analysis
3 Physical handling	2 (2–3)	0.48	0.58	0.81			
16 Equipment/facilities	3 (2–3)	0.54	0.64	0.74			1
2 Intensity	2 (1–2)	0.58	0.65	0.60			
17 Duration	1 (1–2)	0.53	0.61	0.58			
5 Swallow/nutrition	1 (1–1)	0.54	0.63	0.60	(0.36)		
10 Disability	2 (2–3)	0.46	0.56	0.59	(0.30)		2
4 Tracheostomy	2 (1–2)	0.27	0.35	(0.34)	0.40		
11 Discharge planning	3 (2–3)	0.60	0.68	(0.43)	0.52		
6 Communication	2 (2–2)	0.59	0.67		0.59		
15 Medicolegal	2 (2–2)	0.53	0.61		0.74		3
7 Cognitive	1 (1–2)	0.51	0.58		0.67		
8 Behaviour	2 (1–3)	0.36	0.42		(0.44)	0.63	
9 Mood	2 (2–2)	0.39	0.44			0.74	4
1 Medical psych	2 (1–2)	0.44	0.51	(0.45)		0.43	
13 Staff emotional	2 (1–2)	0.62	0.68		(0.56)	0.46	
14 Vocational rehab	1 (1–2)	0.17	0.20			0.45	
12 Family support	2 (1–2)	0.56	0.63	(0.31)	(0.47)	(0.35)	5
Cronbach's alpha			0.86	0.75	0.77	0.71	

*Factor loadings <0.30 are hidden for clarity. **All significant at $p<0.00$.

IQR: interquartile range.

Table III. Patient Categorisation Tool (PCAT): Rasch model summary statistics (overall model fit of the scale)

PCAT Rasch model	Item Location Mean (±95%CI)	Fit residual Mean (±95%CI)	Person Location Mean (±95%CI)	Fit residual Mean (±95%CI)	Item-Trait Interaction χ^2 (DF)/p-value	PSI	Unidimensionality (% Significant t-test)
Analysis 1 (all 17 items)	0.00 (0.08)	−0.02 (0.22)	−0.30 (0.10)	−0.17 (0.08)	378.21 (119)/0.00	0.85	No (16.00)
Locally dependent items combined into 5 subtests							
Analysis 2	0.00 (0.02)	0.46 (0.18)	−0.15 (0.07)	−0.29 (0.09)	56.34 (40)/0.04	0.78	No (7.00)
Locally dependent items combined into 4 subtests							
Analysis 3	0.00 (0.02)	0.27 (0.16)	−0.16 (0.07)	−0.35 (0.09)	31.05 (32)/0.51	0.79	Yes (4.20)
Subtests one and 2 split for DIF by age							
Analysis 4	0.00 (0.02)	0.10 (0.13)	−0.22 (0.07)	−0.36 (0.09)	54.72 (48)/0.23	0.79	Yes (4.20)

DF: degrees of freedom; PSI: Person Separation Index; CI: confidence interval.

alpha was 0.86 for the total scale. Principal components analysis showed that all but 2 items loaded strongly onto the first principal component with loadings ≥0.40, the exceptions being tracheostomy (0.35), and vocational rehabilitation (0.20). These initial analyses were indicative for unidimensionality of the PCAT.

Three components had eigenvalues >1, together explaining 50% of variance. Parallel analysis also indicated 3 factors (see Appendix SII[1]). Varimax rotation revealed 3 factors, but with considerable overlap, which is not unexpected given the evidence for unidimensionality described above (Table II). Given the number of cross-loadings, the extracted components are difficult to interpret meaningfully as separate subscales, but this preliminary investigation outlined item groups that share common variance, which is useful to consider in subsequent Rasch analysis (23) (see below).

Rasch analysis

The likelihood-ratio test indicated the suitability of the Partial Credit Model χ^2 (16)=875.73, $p<0.001$. Table III presents the overall Rasch fit statistics for the sequential analysis of the PCAT and Table IV presents the initial analysis fit statistics for each individual item,

along with the frequency distribution of responses for each of the 3 scoring categories within the 17 items. The initial analysis indicated the overall poor fit to the model marked by significant item-trait interaction and multidimensionality that violated the Rasch model assumptions, but separation reliability was satisfactory (PSI=0.85). Two items (staff emotional and vocational rehabilitation) displayed significant misfit to the model with fit residuals outside of the acceptable range (±2.5) (Table IV). There were no significantly disordered thresholds.

Individual item fit may be affected by local response dependency between items (27). Therefore the item residuals correlation matrix was analysed to identify any influences on responding other than the primary latent trait of interest. The local response dependency identified was between item groups that generally reflected the components identified previously by PCA. Local dependency was addressed using the existing methodology (23, 28) by combining items to accommodate both residual correlations and the PCA findings. Five groups of locally dependent super-items were identified based on a residual correlation matrix that mapped closely to 4 components identified using PCA. Super-items were identified as follows:

Table IV. Rasch model fit statistics and frequency of responses for the 17-item Patient Categorisation Tool (PCAT) (Analysis 1)

Item	Description	Location	Fit residual	χ^2	Score 1	Score 2	Score 3
1	Medical/psychiatric	−1.21	0.98	9.26	37	300	161
2	Intensity	−1.33	−1.45	12.98	60	177	261
3	Physical handling	−1.24	−0.65	6.25	55	216	227
4	Tracheostomy	1.74	−0.93	9.43	465	7	26
5	Swallow/nutrition	0.70	−1.30	6.94	256	173	69
6	Communication*	−0.32	−1.36	9.82	188	113	197
7	Cognitive	−1.17	−0.33	16.23	81	144	273
8	Behaviour	1.04	2.04	30.46	286	171	41
9	Mood	0.40	2.23	19.36	179	249	70
10	Disability	0.16	−0.23	17.88	85	366	47
11	Discharge planning	−0.74	−1.98	23.83	67	298	133
12	Family support	−0.29	−1.20	22.77	76	342	80
13	Staff emotional*	0.80	−2.76	29.85	232	210	56
14	Vocational rehabilitation*	0.26	8.07	122.05	256	132	110
15	Medicolegal	0.43	−1.23	8.46	253	146	99
16	Equipment/facilities	0.24	−0.23	19.11	234	146	118
17	Duration	0.55	−0.05	13.54	184	254	60

*Significant misfit to the Rasch model ($p<0.05$). In this table items are presented in the order in which they appear in the PCAT scale.

- The first 4 items from the first component (Table II): Physical handing, Equipment and facilities, Intensity and Duration.
- The 4 items that cross-loaded on both the first and second components (see Table II): Swallow/nutrition, Disability, Tracheostomy and Discharge planning.
- The 3 remaining items that loaded high only on the second component: Cognition, Communication and Medicolegal.
- The 4 items from the third component: Behaviour, Mood, Medical/psychiatric input and Staff emotional.
- The remaining items: Vocational Rehabilitation and Family support.

Combining the items into 5 super-items improved the overall model fit, but the item-trait interaction χ^2 was still significant and only marginal evidence of unidimensionality was obtained (Table III, Analysis 2). At the individual item level, super-item 5 displayed significant model misfit with a fit residual of 3.20 and local response dependency with super-item 1. Satisfactory model fit was achieved by merging super-items 1 and 5, resulting in the 4 subtests solution (Table III, Analysis 3).

DIF analysis ANOVA indicated significant uniform DIF by age for super-items 1 (F (2, 497)=6.35, p=0.001) and 2 (F (2, 497)=6.62, p=0.001), suggesting that these super-items function differently for adults aged 65 years and older compared with the younger patients with the same level of complexity. Therefore, these were split by age for DIF, resulting in the best overall model fit (Table III, Analysis 4). At this final stage there were no mis-fitting or locally dependent items and item-trait interaction was not significant. Fig. 3 shows the person-item threshold distribution for the final solution indicating very good targeting of the sample by the item thresholds.

Satisfactory fit of the scale to the Rasch model and good coverage of the sample permitted generation of the ordinal-to-interval conversion tables that account

for DIF by age. Both ordinal and interval conversion scores are presented in Appendix I using the original PCAT score range and do not require altering of the scale response format. Using these conversions require a complete set of responses to each of the 17 PCAT items for each patient.

DISCUSSION

This study examined the scaling properties of the PCAT tool in patients with complex rehabilitation needs following acquired brain injury. We used a combination of classical test theory (CTT) and Rasch modelling in the psychometric analysis of a large UK multicentre dataset. Exploratory factor analysis using PCA suggested that the PCAT might reasonably be summed into a single total score reflected in the strong first principal component, but also comprised 3 main factors. In the Rasch analysis, once local response dependency issues were addressed by combining items into 4 testlets or super-items, the model provided good fit for a reliable unidimensional scale, permitted generation of ordinal-to-interval conversion tables that account for DIF by age and increase the accuracy of assessment based on a score range of 16–50.

Measurement models may include either a "reflective" model (where the indicators of a construct are considered to be caused by the construct) or a "formative" model, where the measured variables are considered to be the cause of the construct. The PCAT was derived from a checklist pre-determined by the Department of Health in England to identify patients requiring Level 1 (tertiary) rehabilitation services. Broadly, these are split into "Level 1a" services for profound physical disability, "Level 1c" services for (mainly ambulant) patients with complex cognitive and behavioural needs and Level 1b (mixed) services. The PCAT was designed to capture a diverse range of complex needs across this spectrum. It essentially follows a formative model in which not all items are expected to be correlated, but are individual factors that
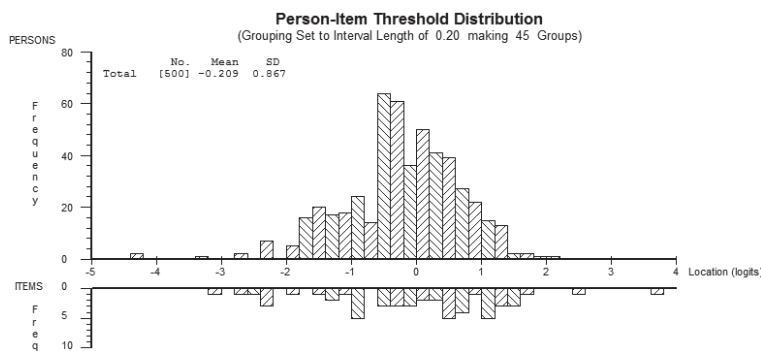

Fig. 3. Person-item threshold distribution of the Patient Categorisation Tool final solution (*n*=500). The person-item threshold distribution for the final solution indicates very good targeting of the sample (persons mean=−0.22) by the item thresholds with no signs of significant flow or ceiling effects.

may contribute to the overall complexity of the patient. Some variables are likely to go together in relatively discrete groupings (for example the requirement for psychiatric input in patients with complex behavioural and psychological needs). Others may run across the component subscales (for example, family support was needed regardless of whether the patient's need were predominantly physical, communicative or cognitive.) Other variables may stand alone; for example, although need for vocational rehabilitation is a valid requirement for intervention from specialist rehabilitation, it stands somewhat alone from items reflecting either profound physical or cognitive dependency. For this reason we did not expect an excellent fit to the Rasch model, although we expected that items may broadly separate into physical and cognitive/behavioural subscales, as indeed they did.

Although the groupings were made on a purely statistical basis, they resonated with clinical practice. For example, whilst one might not intuitively connect complex discharge planning with tracheostomy or swallowing/nutrition problems, the requirement for a tracheostomy or percutaneous endoscopic gastrostomy naturally go with complex disability and patients with all 3 of these issues will inevitably place more burden on discharge planning arrangements. Future clinical developments of the PCAT may therefore include use of the items or super-items in some sort of decision tree to sign-post referral to the most appropriate service for a given individual, weighting of items or the development of a costing algorithm. However, for this study we wished to explore its properties as an interval measure to support future comparison of caseload complexity between different populations and services.

Whilst the PCAT was developed in the UK for identifying patients with complex needs requiring a certain level or type of service, the findings presented here suggest that its measurement properties are consistent with an interval level scale that could perform across a much more graded spectrum of needs. This means that it has potential for application in other countries or health systems with different service structure and design.

One previous study has examined the psychometric properties of the PCAT using CTT techniques in a general neurorehabilitation population (6), but this is the first study to explore its scaling properties using Rasch analysis. The earlier study also provided support for summation of the PCAT items into a single unidimensional scale, although a 2-factor model comprising a "Physical" and "Cognitive/psychosocial" subscale provided the best fit and this is consistent with the clinical diversity described above.

This study has a number of strengths and weaknesses.

Strengths

- The data are drawn from a large dataset comprising data from all the Level 1 and 2 rehabilitation services in England. The results are therefore highly likely to be generalizable within those services.
- The use of both traditional psychometric and Rasch techniques on separate datasets (both of which yield similar results) increases the likelihood that the findings are robust.
- Conversion from ordinal scores into interval-level data did not require modification of the original response format of the PCAT. The conversion permits parametric statistical analysis without violating fundamental assumptions of these tests and also valid comparison with other interval-level data.

Weaknesses

- The data are drawn from a single country (England) and may not be generalizable to other countries. Although the need to identify complexity of needs for rehabilitation is a world-wide requirement, the 3-level structure in the UK is relatively usual. The PCAT is designed to identify patients with highly complex needs requiring tertiary (regional and supra-district) services. This level of selection is not applicable to all countries.
- Because data reporting was optional during the assessment period, we did not expect full data, but only 56% of the episodes had a PCAT recorded. Selection of only those patients with PCAT scores may therefore have introduced some selection bias.
- Currently in England there is much greater service capacity in specialist rehabilitation for patients with complex physical disability than for cognitive behavioural problems. The group with physical disability therefore dominated this particular sample. Further analysis is required to determine whether patients requiring Level 1c services would be better identified with a subset of items comprising the cognitive and behavioural elements (i.e. Parcels 3 and 4) of the PCAT.

Conclusion

The data presented here suggest that the PCAT may reasonably be summed into a single total score, and a transformation table for interval scores has been prepared. However, inspection of the interquartile

ranges of untransformed scores within the clinically assessed categories of needs suggested that raw PCAT scores ≥30 may provide a reasonable indication of Category A needs, 25–29 Category B and 19–24 category C. The added value of transformed scores as a discriminator of category of need (especially in brain-injured patients requiring cognitive behavioural rehabilitation) therefore requires further investigation in future studies.

REFERENCES

1. Specialist rehabilitation for patients with highly complex needs: D02 service specification. NHS England National Programmes of Care 2013 [cited 2014]. Available from: http://www.england.nhs.uk/wp-content/uploads/2014/04/d02-rehab-pat-high-needs-0414.pdf.
2. Wade DT. Complexity, case-mix and rehabilitation: the importance of a holistic model of illness. Clin Rehabil 2011; 25: 387–395.
3. Turner-Stokes L, Williams H, Siegert RJ. The Rehabilitation Complexity Scale version 2: a clinimetric evaluation in patients with severe complex neurodisability. J Neurol Neurosurg Psychiatr 2010; 81: 146–153.
4. The Patient Categorisation Tool (PCAT). UK Rehabilitation Outcomes Collaborative. Clinical Tools 2012 [cited 2017 May 18]. Available from: http://www.kcl.ac.uk/lsm/research/divisions/cicelysaunders/research/studies/uk-roc/tools.aspx.
5. Specialist neuro-rehabilitation services: providing for patients with complex rehabilitation needs. London: British Society of Rehabilitation Medicine. Updated 2015; 2010. Available from: https://www.bsrm.org.uk/publications/publications.
6. Turner-Stokes L, Krägeloh CU, Siegert RJ. The patient categorisation tool: psychometric evaluation of a tool to measure complexity of needs for rehabilitation in a large multicentre dataset from the United Kingdom. Disabil Rehabil 2018 Jan 18: 1–9. [Epub ahead of print].
7. Cano SJ, Barrett LE, Zajicek JP, Hobart JC. Beyond the reach of traditional analyses: using Rasch to evaluate the DASH in people with multiple sclerosis. Multiple Sclerosis 2011; 17: 214–222.
8. Rasch G. Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research). Chicago: The University of Chicago Press; 1960.
9. Bond TG, Fox JC. Applying the Rasch Model. Fundamental measurement in the human sciences. 2nd edn. Toledo: University of Toledo; 2007.
10. Khan A, Chien CW, Brauer SG. Rasch-based scoring offered more precision in differentiating patient groups in measuring upper limb function. J Clin Epidemiol 2013; 66: 681–687.
11. Lundgren Nilsson A, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM) revisited. J Rehabil Med 2011; 43: 884–891.
12. Prodinger B, O'Connor RJ, Stucki G, Tennant A. Establishing score equivalence of the Functional Independence Measure motor scale and the Barthel Index, utilising the International Classification of Functioning, Disability and Health and Rasch measurement theory. J Rehabil Med 2017; 49: 416–422.
13. Siegert RJ, Walkey FH, Turner-Stokes L. An examination of the factor structure of the Beck Depression Inventory-II in a neurorehabilitation inpatient sample. J Int Neuropsychol Soc 2009; 15: 142–147.
14. Siegert RJ, Turner-Stokes L. Psychometric evaluation of the Northwick Park Dependency Scale. J Rehabil Med 2010; 42: 936–943.
15. Turner-Stokes L, Siegert RJ. A comprehensive psychometric evaluation of the UK FIM + FAM. Disabil Rehabil 2013; 35: 1885–1895.
16. O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behav Res Methods Instrument Comput 2000; 32: 396–402.
17. Pett MA, Lackey NR, Sullivan JJ. Making sense of factor analysis: the use of factor analysis for instrument development in health care research. Thousand Oaks, California:

Sage Publications; 2003.

18. Cattell RB. The scree test for the number of factors. Multivariate Behavi Res 1966; 1: 245–276.

19. Horn JL. On subjectivity in factor analysis. Educ Psychol Meas 1967; 27: 811–820.

20. Andrich. Rasch models for measurement. Quantitative Applications in the Social Sciences. London: Sage Publications; 1988.

21. Andrich D. Rasch models for ordered response categories. In: Everitt B, Howell D, editors. Encyclopedia of statistics in behavioral science. New York: John Wiley & Sons; 2005: p. 1698–1707.

22. 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: 1358–1362.

23. Medvedev ON, Siegert RJ, Mohamed AD, Shepherd D, Landhuis E, Krägeloh CU. The Oxford Happiness Questionnaire: transformation from an ordinal to an interval measure using Rasch analysis. J Happiness Stud 2017; 5.

24. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. Br Med J 1995: 310: 170.

25. Smith EVJ. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3: 205–231.

26. Leung YY, Png ME, Conaghan P, Tennant A. A systematic literature review on the application of Rasch analysis in musculoskeletal disease – a special interest group report of OMERACT 11. J Rheumatol 2014; 41: 159–164.

27. Marais I, Andrich D. Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model. J Appl Meas 2008; 9: 105–124.

28. Lundgren-Nilsson A, Jonsdottir IH, Ahlborgjr G, 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.

Appendix I. Ordinal-to-interval Rasch conversion for the modified 18-item Patient Categorisation Tool (PCAT)

Ordinal Score	Interval Age <65 years (95% CI)	Interval Age ≥65 years (95% CI)
16	16.0 (13.5–18.5)	16.0 (13.5–18.5)
17	20.1 (18.4–21.8)	19.2 (19.1–21.0)
18	23.0 (21.8–24.2)	21.5 (20.3–22.6)
19	25.0 (23.9–26.2)	23.0 (21.9–24.2)
20	26.6 (25.6–27.7)	24.2 (23.2–25.2)
21	28.0 (27.0–28.9)	25.2 (24.2–26.1)
22	29.1 (28.2–30.0)	26.0 (25.2–26.9)
23	30.1 (29.2–30.9)	26.8 (25.9–27.6)
24	31.0 (30.2–31.8)	27.4 (26.6–28.2)
25	31.7 (31.0–32.5)	28.0 (27.3–28.8)
26	32.5 (31.7–33.2)	28.6 (27.9–29.3)
27	33.1 (32.4–33.9)	29.1 (28.4–29.9)
28	33.8 (33.1–34.5)	29.6 (28.9–30.4)
29	34.4 (33.7–35.1)	30.1 (29.4–30.8)
30	35.0 (34.3–35.6)	30.6 (29.9–31.3)
31	35.5 (34.9–36.2)	31.1 (30.4–31.8)
32	36.1 (35.4–36.7)	31.6 (30.9–32.3)
33	36.6 (36.0–37.3)	32.1 (31.4–32.8)
34	37.2 (36.5–37.8)	32.6 (31.9–33.3)
35	37.7 (37.0–38.3)	33.1 (32.4–33.8)
36	38.2 (37.6–38.9)	33.6 (32.9–34.3)
37	38.7 (38.1–39.4)	34.1 (33.4–34.7)
38	39.3 (38.6–39.9)	34.5 (33.9–35.2)
39	39.8 (39.1–40.4)	35.0 (34.3–35.7)
40	40.3 (39.6–40.9)	35.5 (34.8–36.1)
41	40.8 (40.2–41.5)	35.9 (35.2–36.6)
42	41.3 (40.7–42.0)	36.4 (35.7–37.1)
43	41.9 (41.2–42.6)	36.9 (36.1–37.6)
44	42.5 (41.7–43.2)	37.4 (36.6–38.1)
45	43.1 (42.3–43.8)	38.0 (37.1–38.8)
46	43.8 (43.0–44.6)	38.7 (37.8–39.6)
47	44.6 (43.7–45.5)	39.8 (38.7–40.9)
48	45.7 (44.6–46.7)	41.5 (40.4–42.6)
49	47.3 (46.3–48.4)	44.6 (40.8–48.4)
50	50.0 (48.9–51.1)	50.0 (46.2–53.8)