Brief Methodological Report

A Rasch Analysis of the Integrated Palliative Care Outcome Scale

Margaret H. Sandham, DClinPsych, Oleg N. Medvedev, PhD, Emma Hedgecock, MHSc (Nursing), Irene J. Higginson, PhD, and Richard J. Siegert, PhD

School of Clinical Sciences (M.H.S.), Auckland University of Technology (AUT), Auckland; School of Public Health and Psychosocial Studies (O.N.M., R.J.S.), Auckland University of Technology (AUT), Auckland; Hospice West Auckland (E.H.), Auckland; and Department of Palliative Care, Policy, and Rehabilitation (I.J.H.), Cicely Saunders Institute, Kings College London, United Kingdom

Abstract

Context. Accurate assessment of a patient's palliative care needs is essential for the timely provision of treatment and support. The Integrated Palliative Care Outcome Scale (IPOS) is an ordinal measure possessing acceptable psychometric properties, but its ability to discriminate precisely between individual symptom levels has not been rigorously investigated.

Objectives. The study aimed to conduct Rasch analysis of the IPOS to evaluate and enhance precision of the instrument.

Methods. Responses of 300 community-dwelling palliative care patients were subjected to Rasch analysis using the partial credit model.

Results. Initial analysis supported the use of the Rasch model and acceptable reliability (person separation index = 0.77) was observed; however, unsatisfactory model fit was found. Local dependency between items was resolved through the creation of super-items, which increased model fit, reliability (person separation index = 0.80), and unidimensionality. There were no misfitting super-items or differential item functioning by age, rater, sex, or ethnicity. The IPOS showed satisfactory coverage of symptoms within the present clinical sample, with the ability to assess higher severity patients.

Conclusion. The modified IPOS showed excellent reliability for a clinical measure in assessing the overall palliative care needs of a patient. The provided ordinal-to-interval conversion table accounts for unique contribution of each symptom to the overall symptom burden and easy to use without the need to modify the original IPOS format. J Pain Symptom Manage 2019;57:290–296.

Key Words

Integrated Palliative Care Outcome Scale (IPOS), Rasch analysis, psychometrics, assessment, validation

Introduction

Accurate assessment of patient's palliative care needs is essential. The use of validated, standardized assessment measures in palliative care enables the early identification and response to changes in patients' status and accurate evaluation of the efficacy of interventions. The Integrated Palliative Care Outcome Scale (IPOS) is a brief patient-reported or clinician-rated measure used to assess a patient's palliative care needs, including physical and psychological symptoms, and support needs. The IPOS is widely used in palliative care and was developed from integrating the original Palliative Outcomes Scale (POS) as developed by Hearn and Higginson,[1] and the POS symptom list. The POS has been widely adopted in palliative care, with use in 14 countries reported.[2] The POS has undergone psychometric evaluation using traditional methods such as reliability,[1] exploratory factor analysis,[3,4] and confirmatory factor analysis.[3] Rasch analysis has been used with selected

Address correspondence to: Margaret H. Sandham, DClinPsych, School of Clinical Sciences, Auckland University of Technology (AUT), Auckland, New Zealand. E-mail: Margaret.sandham@aut.ac.nz

Accepted for publication: November 18, 2018.

0885-3924/$ - see front matter
https://doi.org/10.1016/j.jpainsymman.2018.11.019

POS items to develop a palliative care health classification system and supported its internal validity[5]; however, as yet, Rasch analysis has not been used with the IPOS and conversion tables have not been produced. Cognitive interviews evaluating the acceptability and utility of the IPOS showed that the questionnaire was easy to comprehend, acceptable, and not burdensome to complete.[6] Further psychometric studies on the IPOS are needed to establish its reliability, clinical utility, and validity.

Rasch analysis[7,8] is a robust psychometric method that has been widely applied to evaluate and enhance the measurement properties of outcome measures. Rasch analysis offers important advantages over classical test theory approaches such as ensuring the measure is working well for different population groups (measurement invariance) (e.g., age, sex) and measures one specific domain (the overall symptom burden) without bias toward other unrelated domains(e.g., personality).[9] Rasch analysis includes scrutiny at the individual item level and assessment of differential item functioning (DIF), which allow a researcher to identify and potentially eliminate specific sources of item bias leading to improvement of both reliability and validity of a measure,[9,10] through means such as removing items. When satisfactory fit to the Rasch measurement model is achieved, patients can be ordered on an interval-level scale based on their level of the underlying construct being measured.[10] Interval level measurement ensures that the distance between points on the rating scale (e.g., 1-2-3) is equal, rather than a comparative ranking with unequal distances between points. This study aims to undertake Rasch analysis to assess the psychometric properties of the IPOS and to develop a conversion table to convert ordinal raw scores into interval-level scores to enhance the precision of the scale.

Methods

Participants

Data were collected as part of the routine care of patients enrolled in a community hospice setting between December 2017 and August 2018. Patients were administered the study measures if they were aged over 18 years and able to complete the study measures independently, with help, or by proxy. Exclusion criteria were inadequate English language to complete the IPOS if no translator available or if the clinical team judged them as being too unwell or distressed to complete the measures. A total of 300 patients were evaluated (females $n = 133$, males $n = 167$), ages ranging from 35 to 98 years, mean = 68 years, SD = 13.9. The sample size satisfied the requirements for the planned Rasch analysis.[11,12] Age groups were created based on frequency distribution to ensure equal group sizes, and ethnicity groups were created based on New Zealand Census categories. Three age groups were created for evaluating DIF: <60 years $n = 98$; 61−75 years $n = 106$; >76 years $n = 96$. Ethnicities represented were Caucasian, $n = 183$; Maori, $n = 37$; Pacific Island, $n = 28$; Asian, $n = 22$; Other, $n = 30$. Severity as measured by the Australian Modified Karnofsky Performance Scale (AKPS)[13] ($n = 267$) was mean = 5.88, SD = 1.65, mode 60% (interquartile range lower quartile score 50%, middle quartile score 60%, upper quartile score 70%). Participants predominantly had cancer (87.4%), with the remainder experiencing nonmalignant terminal illness (12.6%). Questionnaires were completed on admission; 44 (18%) participants provided their responses without assistance; 142 (57%) got help from relatives; and the remaining 40 (16%) with help from a clinician. DIF by diagnosis could not be evaluated owing to small numbers within each diagnostic group, and comparison between cancer and noncancer diagnostic groups significantly differed by size ($P < 0.001$).

Instruments

Study Measures: 1) The IPOS[1] includes 17 items measuring palliative care needs in the domains of physical and psychosocial functioning. Symptoms are assessed on a 0−4 Likert scale, and two open-ended questions have free response so that patients may record symptoms of importance to them. The final item asks if the IPOS was completed independently, with help, or by a proxy. The AKPS[13] is a brief clinician-administered rating scale based on observations of the patient's performance on three dimensions of work, activity, and self-care. Patients are rated from 0% (dead) to 100% (normal, no complaints, no evidence of disease). Clear criteria are provided for increments of 10%. The AKPS has demonstrated good psychometric properties in Australian palliative care populations.[13]

Procedures

All participants gave signed informed consent to participate in assessment and to use their health information for audit purposes, including data analysis. Verbal informed consent was gained by the clinician to complete the questionnaire.

Statistical Analysis

Rasch analysis was undertaken using RUMM2030.[14] Analysis was conducted following the procedure described in Medvedev, Turner-Stokes[15] starting with

assessment of the overall and individual item fit, and examining local dependency between items. In the present analysis, deleting items was considered as the last resort to improve Rasch model fit. The initial analysis involved evaluation of the overall and individual items fit to the Rasch model and examination of residual correlations between items because it influences the overall goodness of fit.[16] If local dependency was found between items (indicated by the magnitude of the residual correlations, e.g., 0.20), locally dependent items were meaningfully combined into super-items.[16] For example, nausea and vomiting could be conceptually combined into a super-item because they represent a related physiological experience.

The following Rasch model fit criteria were used in the present analysis[17,18]: First, the overall Rasch model fit should be supported by a nonsignificant item-trait interaction chi-square ($P > 0.05$, Bonferroni adjusted). Second, ideal fit was that overall fit residuals for item and person are expected to have a mean close to 0.00 and an SD close to 1.00. Third, individual item fit residuals should be within the range +/−2.50. Fourth, a between-sample subgroups ANOVA test should indicate that no items display DIF across groups (e.g., males vs. females). Finally, reliability will be assessed using the person separation index (PSI) in Rasch analysis, the value of which is interpreted similar to Cronbach's alpha.

Dimensionality was examined by using a form of factor analysis, that involved evaluating the first principal component of the residuals after removing the latent trait.[19] Person locations of the two item groups (one with the highest positive and the other with the lowest negative loadings on the first principal component of the residuals) were compared by an independent t-test. Unidimensionality was examined using the standardized unweighted item fit residuals. Evidence of unidimensionality is obtained if the percentage of significant t-test was <5% or if the lower bound surrounding the number of significant t-tests overlapped 5%.[18] When the acceptable Rasch model fit is achieved, the person-item thresholds distribution is assessed to evaluate coverage of the sample symptom levels by items thresholds. Finally, the conversion tables are produced using person estimates of the Rasch model to transform ordinal raw scores into interval-level data to improve accuracy of measurement.[15]

Results

Data Preparation

There was 3.3% of data missing. Correction for missing data was not undertaken as the Rasch model is robust with missing data,[20] and data were missing completely at random. Nonordinal Items Q1 (1−3) and Q2 (1−3) were removed as they were unsuitable for the Rasch measurement model. Q10 was removed as it was a demographic item.

Initial Analysis

A likelihood ratio test was first conducted to confirm the appropriateness of the partial credit Rasch model[7] for the current data set, which was supported by rejecting the rating scale model X^2 (47) = 259.42, $P < 0.0001$. Table 1 summarizes the overall fit statistics for the Rasch models. Initial analysis showed unsatisfactory overall model fit (X^2 (136) = 202.86) but reliability (>0.70) was acceptable for group assessment[18](e.g., comparison of group as a whole rather than an individual) as indicated by the PSI of 0.77.[18] Examination of individual item thresholds was conducted, and no significantly disordered thresholds were found.

Table 2 presents individual item fit statistics and shows that Items 15, 16, and 17 have slightly elevated fit residuals above the fit criteria of 2.5. At this stage, no evidence for unidimensionality was found as indicated by 15% of significant t-tests.

Second Analysis With Super-Items

The overall model fit and dimensionality may be influenced by local dependency between items, so we evaluated a residual correlation matrix. Local dependency was found between most of the items, indicated by residual correlation at the level of 0.20 and higher. Therefore, locally dependent items were combined into seven super-items based on their residual correlations (see Table 2). Local dependency may result from response dependency (e.g., weakness and problem solving because if someone feels weak, they may also perceive themselves to be less capable to solve problems) rather than conceptual relationships between items.[12,21] However, in instances where high residual correlations were found between several items, conceptually meaningful pairing was

Table 1
Summary of Overall Rasch Model Fit Statistics

	Item Residual		Person Residual		Goodness of Fit		PSI	Independent t-Test	
Analysis	Mean	SD	Mean	SD	X^2 (df)	*P*	Value	%	95% CI
1	0.51	1.30	−0.08	1.10	202.86 (136)	<0.001	0.77	15.33	12.87
2	0.30	0.33	−0.21	0.95	56.18 (56)	0.468	0.80	5.00	2.30

PSI = person separation index.

Table 2
Item Fit Statistics of the Rasch Model for Initial and Final Analyses

Items	Item Description	Location	SE	FitResid	ChiSq
1	Pain	−0.22	0.06	1.70	4.77
2	Shortness of breath	−0.01	0.05	0.91	2.41
3	Weakness or lack of energy	−0.77	0.06	−0.30	9.51
4	Nausea	0.86	0.07	−0.92	7.62
5	Vomiting	1.79	0.08	−0.65	6.67
6	Poor appetite	−0.23	0.05	0.10	9.98
7	Constipation	0.12	0.06	−0.36	4.14
8	Sore or dry mouth	0.24	0.06	−0.06	4.70
9	Drowsiness	0.66	0.06	−0.62	16.53
10	Poor mobility	−0.40	0.05	1.31	4.43
11	Anxiety or worry about illness or treatment	−0.13	0.05	−1.28	12.44
12	Family or friends anxiety	−0.80	0.05	0.29	1.09
13	Depression	0.31	0.06	−0.21	7.68
14	Feeling at peace	−0.31	0.05	0.73	2.19
15	Sharing with family	−0.45	0.05	2.55*	11.32
16	Sufficiency of information	−0.47	0.04	2.98*	19.23
17	Practical problem solving	−0.21	0.05	2.51*	16.32
Super-items					
S1	Q4 (nausea) + Q5 (vomiting)	0.49	0.04	0.32	8.84
S2	Q6 (appetite) + Q9 (drowsiness) +Q16 (information)	0.07	0.03	0.22	9.31
S3	Q2 (SOB)+Q7 (constipation) + Q15 (sharing with family)	−0.02	0.03	0.25	4.63
S4	Q8 (dry mouth) + Q11 (anxiety)	0.01	0.04	−1.22	15.84
S5	Q10 (mobility) + Q12 (family worry)	−0.48	0.04	0.56	6.70
S6	Q1 (pain) + Q13 (depression) + Q14 (peace)	0.25	0.03	0.94	6.35
S7	Q3 (weakness) + Q13 (problem solving)	−0.33	0.04	1.06	4.51

*Significant misfit to the Rasch model ($P < 0.05$).

prioritized. For example, anxiety and dry mouth are both symptoms of anxiety; these were combined. This combination resulted in the best model fit (see Table 1, Analysis 2), improved reliability (PSI = 0.80), and unidimensionality. Therefore, Rasch analysis provided evidence that the scale measures the overall symptom burden affecting the patient as the overarching latent construct. There were no misfitting super-items (Table 2), or differential super-item functioning by ethnicity, age, sex, and whether the questionnaire was completed by the patient or a proxy. DIF by AKPS scores was not completed owing to an uneven distribution of scores across the range of the AKPS.

Figure 1 shows the person-item threshold distribution indicating that 95% of individual symptom levels (abilities) are satisfactorily covered by item thresholds of IPOS. The item thresholds on the right-hand (more difficult) side of the scale cover substantially higher levels of symptom severity not present in the current sample.

Ordinal-to-Interval Conversion

The recently published guidelines for reporting Rasch analysis require the production of ordinal-to-interval conversion tables derived from the person estimates of the Rasch model.[22] The use of conversion tables enhances the ability of the measure to discriminate between individual patients on the overall symptom burden using the IPOS. An algorithm was developed to convert ordinal scores to interval-level data using the original response format of the scale (see Table 3). This table includes ordinal raw scores calculated by adding individual scores of the original IPOS Items Q2 (symptoms), excluding the free response symptoms, Q3 to Q9. Questions excluded from the raw score conversion table were Q1, free response items from Q2, and Q10. After the raw score is calculated, researchers or clinicians can find a corresponding interval-level score in both logit units (original Rasch estimates) or in a scale metric.

Discussion

The present study is the first Rasch analysis of the IPOS, which is a widely used palliative care outcome measure. We have demonstrated that the IPOS satisfactory meets the expectations of the unidimensional Rasch model including invariance across population groups, unidimensionality, and consistent interval measurement units along the continuum of the scale. Although the scale appears to measure different and unrelated symptoms (e.g., psychological symptoms and physical symptoms), the unidimensionality found in this study is suggestive that after Rasch transformation of scores, the scale measures the overall construct of palliative care needs. This can be illustrated by an example of extracting a vitamin C from different fruits such as oranges, lemons, and apples. These are different fruits, but they all contain the same element, vitamin C to a different degree. This can be compared to item location or difficulty in Rasch analysis. Some symptoms contribute more to the overall palliative care needs (e.g., vomiting has the highest item location or difficulty), whereas other symptoms have

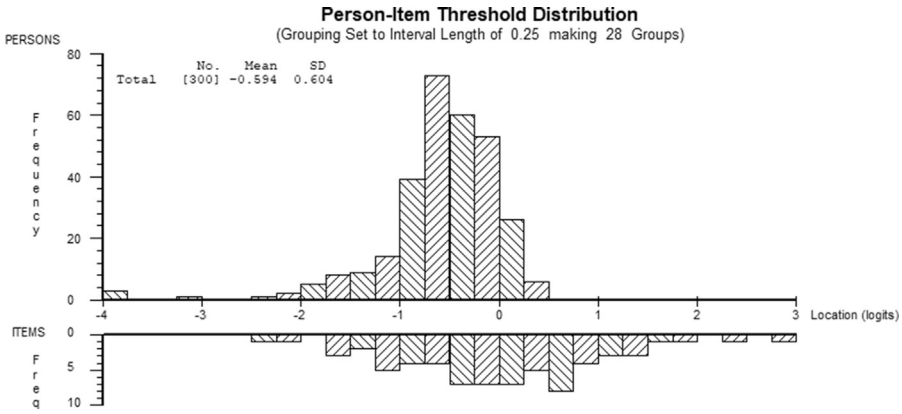

Fig. 1. Person-item threshold distribution.

only a minor contribution (e.g., anxiety experienced by family and friends about the patient). A further advantage of these results is that the ordinal scale scores can be transformed into interval-level data and analyzed with parametric statistics lending greater statistical power and precision.[15] The ordinal-to-interval conversion tables published here provide researchers with robust metrics to evaluate outcomes using the IPOS and to make valid statistical inferences with bio physiological and neuropsychological interval data such as biomarkers, EEG, and heart rate.

Table 3
Ordinal to Interval Conversion Table

Ordinal	Interval		Ordinal	Interval	
Raw Score	Logit	Scale	Raw Score	Logit	Scale
0	−3.98	0.00	35	0.06	33.54
1	−3.21	6.44	36	0.10	33.89
2	−2.70	10.65	37	0.14	34.23
3	−2.37	13.40	38	0.18	34.58
4	−2.12	15.44	39	0.22	34.93
5	−1.93	17.06	40	0.27	35.28
6	−1.77	18.40	41	0.31	35.63
7	−1.63	19.56	42	0.35	35.99
8	−1.51	20.55	43	0.40	36.36
9	−1.40	21.45	44	0.44	36.73
10	−1.30	22.26	45	0.49	37.10
11	−1.21	23.00	46	0.53	37.49
12	−1.13	23.67	47	0.58	37.90
13	−1.06	24.30	48	0.63	38.32
14	−0.98	24.90	49	0.68	38.75
15	−0.92	25.46	50	0.74	39.20
16	−0.85	25.99	51	0.79	39.67
17	−0.79	26.49	52	0.85	40.16
18	−0.74	26.97	53	0.92	40.69
19	−0.68	27.43	54	0.98	41.25
20	−0.63	27.88	55	1.06	41.85
21	−0.57	28.31	56	1.13	42.49
22	−0.52	28.74	57	1.22	43.19
23	−0.47	29.14	58	1.31	43.97
24	−0.43	29.54	59	1.41	44.82
25	−0.38	29.93	60	1.53	45.78
26	−0.33	30.32	61	1.66	46.89
27	−0.29	30.69	62	1.82	48.16
28	−0.24	31.06	63	2.00	49.66
29	−0.20	31.43	64	2.22	51.48
30	−0.16	31.79	65	2.49	53.76
31	−0.11	32.14	66	2.85	56.78
32	−0.07	32.50	67	3.40	61.29
33	−0.03	32.85	68	4.20	68.00
34	0.02	33.20			

The present study undertaken in community hospice patients supports the use of the IPOS total score in this population. We used the formation of clinically meaningful super-items to improve fit and reliability, and local dependency was eliminated. Therefore, IPOS items did not show DIF by age, sex, rater, or ethnicity. Of note, the previously reported issues with item misfit and DIF on the earlier developed POS identified in Rasch analysis[5] were not present in the IPOS with the current sample, suggesting that the refinement of the POS to create the IPOS improved psychometric properties of the measure.

The range of symptom severity assessed by the IPOS captured over 95% of the present sample and showed that it will be able to assess more severe patients in the population. The IPOS was less able to assess patients with few mild symptoms well; however, patients with mild symptoms are not the target of the IPOS. In these instances, broader psychological and quality of life measures such as the World Health Organisation Quality of Life[23] can be used. This distribution of item difficulty, with the ability to capture more severe patients, is advantageous for a clinical scale such as the IPOS. The most difficult items were nausea and vomiting, suggesting that presence of these symptoms is associated with the overall high symptomatic severity. The easiest items to endorse were the mobility and family worry items, indicating that most participants have decreased mobility and are aware of their family and friends being worried about them.

A strength of this study is that it evaluated routine clinical data from hospice patients. This supports the generalizability to other community hospice settings; however, there may be an underrepresentation of higher severity patients. Despite this, our results suggested that the IPOS is capable accurately assess patients experiencing greater symptoms severity. Evaluation of the clinical utility of transformed scores and responsivity to clinical change is an area for future research.

Limitations

Data were predominantly collected at admission to a community hospice service, consequently there may be an underrepresentation of higher severity patients such as from inpatient units, or from those that clinicians felt were too unwell to complete the measure. Ethnic groups were not evenly represented for DIF analysis; however, this is a reflection of the natural population in New Zealand. DIF by ethnicity should be evaluated using data from the population where it will be used, as this sample was predominantly Caucasian. We did not evaluate DIF by diagnosis as there was a significant difference between the sizes of the cancer and non-cancer groups, and the individual diagnostic categories were not large enough to evaluate separately. However, the Rasch model is sufficiently robust suggesting that if overall model fit is achieved, then the individual differences are less problematic and the results are generalizable across sample population.[9]

Conclusions

This is the first Rasch analysis of the IPOS, and the psychometric properties found support the use of the IPOS as a clinical and research measure. The IPOS with super-items demonstrated the best Rasch model fit, unidimensionality, invariance across sample groups (no DIF), and excellent reliability for a clinical measure, that permits accurate assessment of symptoms at the individual level. The coverage of symptom severity of the IPOS was adequate for a clinical sample, with room for assessment of more severe symptoms compared to our sample with a slight overrepresentation of the lower severity patients. We have produced and published the conversion table to convert the total ordinal scores into interval-level data based on person estimates of the Rasch model.

Disclosures and Acknowledgments

The researchers wish to thank the team at Hospice West Auckland and the New Zealand Palliative Care Outcome Measurement Group for providing clinical support and guidance to collect these data.

Ethical approval was granted by the authors' institutional Ethics Committee (AUTEC; 18/98), and the internal ethics board at the community hospice where the research was undertaken. This research did not receive any specific grant from the funding agencies in the public, commercial or not-for-profit sectors.

References

1. Hearn J, Higginson IJ. Development and validation of a core outcome measure for palliative care: the palliative care outcome scale. Palliative Care Core Audit Project Advisory Group. Qual Health Care 1999;8:219–227.

2. Bausewein C, Le Grice C, Simon S, et al. The use of two common palliative outcome measures in clinical care and research: a systematic review of POS and STAS. Palliat Med 2011;25:304–313.

3. Siegert RJ, Gao W, Walkey FH, et al. Psychological well-being and quality of care: a factor-analytic examination of the palliative care outcome scale. J Pain Symptom Manage 2010;40:67–74.

4. Higginson IJ, Donaldson N. Relationship between three palliative care outcome scales. Health Qual Life Outcomes 2004;2:68.

5. Dzingina M, Higginson IJ, McCrone P, et al. Development of a patient-reported palliative care-specific health classification system: the POS-E. Patient-Patient-Centered Outcomes Res 2017;10:353–365.

6. Schildmann EK, Groeneveld EI, Denzel J, et al. Discovering the hidden benefits of cognitive interviewing in two languages: the first phase of a validation study of the Integrated Palliative care Outcome Scale. Palliat Med 2016;30:599–610.

7. Masters G. Rasch model for partial credit scoring. Psychometrica 1982;47:149–174.

8. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press, 1960.

9. Hobart JC, Cano SJ, Zajicek JP, et al. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol 2007;6:1094–1105.

10. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. Int J Nurs Stud 2009;46:380–393.

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

12. Lundgren-Nilsson A, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM) revisited. J Rehabil Med 2011;43:884–891.

13. Abernethy AP, Shelby-James T, Fazekas BS, et al. The Australia-modified Karnofsky Performance Status (AKPS) scale: a revised scale for contemporary palliative care clinical practice. BMC Palliat Care 2005;4:7.

14. RUMM Laboratory Pty Ltd. RUMM 2030 Software Application 2010.

15. Medvedev O, Turner-Stokes L, Ashford S, et al. Rasch analysis of the UK Functional Assessment Measure in patients with complex disability after stroke. J Rehabil Med 2018;50:420–428.

16. Christensen KB, Kreiner S, Mesbah M. Rasch models in health. John Wiley & Sons, 2013.

17. Gustafsson JE. Testing and obtaining fit of data to the Rasch model. Br J Math Stat Psychol 1980;33:205–233.

18. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what it is 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.

19. Smith EJ. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal

component analysis of residuals. J Appl Meas 2002;3: 205–231.

20. Boone W, Staver J, Yale M. Rasch Analysis in the Human Sciences. New York, USA: Springer, 2014.

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

22. Leung Y, Png ME, Conaghan P, et al. 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.

23. WHOQOL Group. The World Health Organization quality of life assessment (WHOQOL): Development and general psychometric properties. Social Sci Med 1998;46: 1569–1585.